THE UNIVERSITY OF
SYDNEY

Resource Allocation and Feedback in Wireless Multiuser Networks

Malcolm Egan

A THESIS SUBMITTED FOR THE DEGREE OF

Doctor of Philosophy

Faculty of Engineering and Information Technologies
THE UNIVERSITY OF SYDNEY

2014

Abstract

THIS thesis focuses on the design of algorithms for resource allocation and feedback in wireless multiuser and heterogeneous networks. In particular, three key design challenges expected to have a major impact on future wireless networks are considered: cross-layer scheduling; structured quantization codebook design for multiuser multiple-input multiple-output (MU-MIMO) networks with limited feedback; and resource allocation to provide physical layer security.

The first design challenge is cross-layer scheduling, where policies are proposed for two network architectures: user scheduling in single-cell multiuser networks aided by a relay; and base station (BS) scheduling in coordinated multipoint (CoMP). These scheduling policies are then analyzed to guarantee satisfaction of three performance metrics: symbol error probability (SEP); packet delay; and packet loss probability (PLP) due to buffer overflow. The concept of the τ-achievable PLP region is also introduced to explicitly describe the tradeoff in PLP between different users.

The second design challenge is structured quantization codebook design in wireless networks with limited feedback, for both MU-MIMO and CoMP. In the MU-MIMO network, two codebook constructions are proposed, which are based on structured transformations of a base codebook. In particular, the first construction is based on the Householder transform, and the second construction is based on group representation theory. Both constructions are shown to reduce sum-rate loss due to rank-deficient quantized channel matrices when zero-forcing precoding is employed. A rate-outage lower bound is also derived and used to construct optimal base codebooks, which are shown to have the structure of an important class of equiangular tight frames. In the CoMP network, a low-complexity construction is proposed to solve the problem of variable codebook di-

mensions due to changes in the number of coordinated BSs. The proposed construction is shown to have comparable performance with the standard approach based on a random search, while only requiring linear instead of exponential complexity.

The final design challenge is resource allocation for physical layer security in MU-MIMO. To guarantee physical layer security, the achievable secrecy sum-rate is explicitly derived for the regularized channel inversion (RCI) precoder. The optimal regularization parameter is then derived in the large-system regime with equal power allocation. To improve performance, power allocation and precoder design are jointly optimized using a new algorithm based on convex optimization techniques. Simulations show that the proposed algorithm outperforms equal power allocation by up to 20%.

Statement of Originality

The work presented in this thesis is the result of original research carried out by myself, in collaboration with my supervisor Prof. Iain B. Collings, while enrolled in the School of Electrical and Information Engineering at the University of Sydney as a candidate for the Doctorate of Philosophy. It has not been submitted for any other degree or award in any other university or educational institution.

Malcolm Egan, 2014

Acknowledgements

First and foremost I would like to thank my supervisor Iain Collings for his encouragement, patience, and advice throughout my entire candidature. This thesis would not have been possible without his support, technical expertise and willingness to share the knowledge he has gained from his wide experience in research.

I would also like to thank my collaborators on projects: CK Sung, Wei Ni, Phil Yeoh, Maged Elkashlan, Giovanni Geraci, Jinhong Yuan, Adeel Razi, Siavash Bayat, Ido Nevat, and Gareth Peters. Their friendship, advice, and input have greatly improved the content and presentation. I am also grateful for the exposure to new ideas and methods that I would not have come across otherwise.

In Sydney, I have greatly appreciated the discussions over lunches and coffee at CSIRO, the University of Sydney, and UNSW. In particular, I would like to thank Shuang Tian, Jun Li, Henry Chen, Qimin You, Raymond Louie, Wibowo Hardjawana, Tao Yang, Jonas Yang, Yiwen Li, Seh Chun Ng, David Qi, Zijie Zhang, Joao Rebellato, and Zhuo Chen.

I want to thank Vaughan Clarkson for his continued support since my time at UQ. I have also always enjoyed the fun discussions at AusCTW; particularly with Robby McKilliam, Vaughan, Jinhong, Maged, and Phil. I am also appreciative of the discussions with Michael Honig during his visit to CSIRO.

Thanks are also due to CSIRO, where I was located for much of my PhD. I acknowledge the funding for research and travel throughout my candidature. I would also like to thank Fiona McKeown for her help with administrative matters during my time at CSIRO.

During my studies, I was fortunate enough to travel to several international conferences and visit overseas institutions. I would like to thank the members of WCNG at the

University of Texas at Austin for their hospitality while I was visiting. In particular, I owe thanks to Robert Heath, Salam Akoum, Amin Khalek, Omar El Ayach, Ken Pesyna, Jonathan Starr, and Kien Truong. I also owe thanks to Gareth Peters for welcoming me at UCL, and Martijn Arts, Anke Schmeink, and Rudolf Mathar at RWTH Aachen.

My studies have largely been supported by scholarships from the Australian Government and the University of Sydney. In particular, I acknowledge the support given in the form of the Australian Postgraduate Award, Vice-Chancellor's Research Scholarship, and the Norman I. Price Scholarship.

Thanks are also due to my Sydney friends Emilie, Callum, Jimmy, Simone, and Terry. These years in Sydney would have been much less enjoyable with you all.

Last, but certainly not least, I want to thank all my family for the huge amount of support throughout all my studies. I am overwhelmingly grateful for the accommodation and fun times in Washington DC by Jen and Alastair, and in London by Edwina, Harriet and Tom, and Alice. I am also thankful for the kind words of support of my grandparents. To my parents and Andrew, no words of thanks are enough for all the encouragement and support you have given me.

Contents

List of Figures

List of Tables

Glossary

Acronym	Description
AaF	Amplify and forward
AWGN	Additive white Gaussian noise
BE	Best effort
BPSK	Binary phase-shift keying
BS	Base station
CCDF	Complementary cumulative density function
CDF	Cumulative distribution function
CDMA	Code division multiple access
CI	Channel inversion
CoMP	Coordinated multipoint
CSI	Channel state information
CU	Central unit
DPS	Delay in packet scheduling
ESC	Expected square correlation
ETF	Equiangular tight frame
FDD	Frequency division duplex
FFR	Fractional frequency reuse
FIFO	First-in first-out
GR	Group representation
HOL	Head of line

Hz	Hertz
i.n.d	Independent non-identically distributed
KKT	Karush-Kuhn-Tucker
ISP	Internet service provider
LTE	Long term evolution
MIMO	Multiple-input multiple output
MISO	Multiple-input single output
MISOME	Multiple-input single-output eavesdropper
MMSE	Minimum mean square error
MSE	Mean square error
MU-MIMO	Multi-user MIMO
NUM	Network utility maximization
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
PAM	Pulse amplitude modulation
PDF	Probability density function
PLP	Packet loss probability
PMF	Probability mass function
QPSK	Quadrature phase-shift keying
RCI	Regularized channel inversion
RCI-EP	Regularized channel inversion with equal power allocation
RT	Real time
s	Seconds
SEP	Symbol error probability
SFR	Soft frequency reuse

SINR	Signal-to-interference and noise ratio
SNR	Signal-to-noise ratio
WMRN	Wireless multimedia relay network
ZF	Zero-forcing

*

Notation

Symbol	**Description**.
j	Imaginary unit, $\sqrt{-1}$.
$\mathbb{R}, \mathbb{C}, \mathbb{Z}_k$	Real numbers complex numbers, integers mod k.
$\mathbb{R}^n$	n-dimensional real vector space.
$\mathbb{C}^n$	n-dimensional complex vector space.
$\mathbb{C}P^{n-1}$	n-dimensional complex projective space.
$\mathcal{G}(n,k)$	Grassmannian manifold of k-dimensional subspaces in $\mathbb{C}^n$.
$\mathrm{tr}(\mathbf{X})$	Trace of the matrix $\mathbf{X}$.
$\det(\mathbf{X})$	Determinant of the matrix $\mathbf{X}$.
$\otimes$	Kronecker product.
$\oplus$	Direct sum.
$\mathbf{I}_{n\times n}$	$n \times n$ identity matrix.
$\mathbf{0}_{n\times m}$	$n \times m$ matrix of zeros.
$(\cdot)^T$, $(\cdot)^\dagger$	Matrix transpose, complex conjugate.
$\mathrm{diag}(a_1,\ldots,a_n)$	$n \times n$ diagonal matrix with $a_1,\ldots,a_n$ on the diagonal.
$\mathbf{X} \succeq 0$	non-negative definite matrix.
$\mathbf{x} \succeq \mathbf{0}$	Each component of $\mathbf{x}$ satisfies $x_i \geq 0$.
$\|\mathbf{x}\|$	Euclidean norm of the vector $\mathbf{x}$.
$\|\mathbf{X}\|_F$	Frobenius norm of the matrix $\mathbf{X}$.

$\Pr(\cdot)$	Probability.
$\mathrm{E}[\cdot]$	Expectation.
$\mathcal{CN}(\mu, \sigma^2)$	Circularly symmetric complex normal distribution with mean μ and variance σ^2.

Chapter 1
Introduction

1.1 Heterogeneous Networks and Small-Cells

In the last decade a major paradigm shift has occurred: wireless multiuser cellular networks are now dominated by data and video traffic. Driven by the shift, wireless operators are seeking to offer services with a significant reduction in dollar per bit cost. This has led to a fundamental re-think of the wireless access network, resulting in heterogeneous network architectures and small-cells.

Traditionally, wireless multiuser networks consisted only of large, high power, high cost base stations (BSs) and a large number of associated users. While the traditional setup easily supported voice calls with data rates on the order of 10 kb/s [Chandrasekhar et al., 2008], it fails to support modern devices with high-rate video and data applications.

Put simply, BSs and users need to be closer together in order to support modern capacity-hungry devices. The key to reducing the distance between BSs and users is high BS spatial density. Unfortunately, increasing the spatial density of traditional BSs–known as macrocell BSs–is infeasible due to high infrastructure costs and severe intercell interference [Chandrasekhar et al., 2008, Andrews et al., 2012, Ghosh et al., 2012, Damnjanovic et al., 2011].

The proposed new architecture to achieve the goal of high BS spatial density is based on small-cells. These small, low power, low cost BSs are overlaid on the macrocell network to achieve the goal of high spatial BS density without the financial cost associated with networks consisting of only macrocell BSs. This small-cell based architecture is already being implemented to a limited degree in the case of femtocells [Chandrasekhar

et al., 2008] and in some commercial cellular deployments; for example, SK Telecom in Korea [Small Cell Forum, 2012]. The result of the new architecture is a fundamental change in the structure of the cellular network: *heterogeneity*; due to variations in the received signal-to-interference and noise ratio (SINR) and the number of users each BS can support.

Although effective interference mitigation in wireless multiuser networks with heterogeneous architectures is feasible, there remain difficult challenges to design practical schemes and effectively manage network resources, which are distributed over the network. These challenges mean that naive implementations–without appropriate interference mitigation and resource allocation–in fact lead to a lower system capacity than traditional networks. To achieve the potential of heterogeneous wireless multiuser networks with both macrocell and small-cells, effective network organization is the key.

1.2 Network Organization

The basic design challenge in any communication network is to ensure that data is received by the intended users reliably and on-time. In general, and especially in wireless multiuser networks, this challenge requires the organization of a massive number of distributed BSs and users. Optimal network organization is therefore a confronting optimization problem over space, time and frequency. Owing to the intractability of this optimization problem, the network organization is relaxed into several design challenges, which aim to offer high performance while only requiring the implementation of low complexity algorithms.

Three fundamental design challenges are essential to wireless multiuser network organization:

1. user association;
2. resource allocation;
3. and an interference mitigation strategy.

Importantly, each of these three design challenges is coupled to the others. In fact, the

high spatial BS density and resulting interference means that there is a particularly strong coupling between each design challenge compared with traditional macrocell-only networks.

1.3 User Association

The network decision for the BS or BSs that each user should be serviced by is known as user association. Although the design challenge of user association has been present since the very early wireless cellular networks, new approaches have been motivated by wireless multiuser networks with small-cell based architectures and cooperating macrocell BSs. This is due to an increase the BS density, and also variations in transmit power levels and the number of users that can be supported per BS.

In traditional cellular networks, users are typically associated to the BS with the highest received power [Andrews et al., 2012]; an effective technique only in networks where each BS has the same transmit power and hence inappropriate in heterogeneous networks [Ye et al., 2013, Andrews, 2013]. In contrast, for interference-limited networks–particularly heterogeneous networks with small cells–the main factor in determining each user's rate is the number of users associated to the same BS (the BS loading). As macrocell BSs are heavily loaded with high transmit power, load balancing is necessary; in particular, users can be associated to a small-cell instead of a macrocell BS, even when the received power is lower (known as range expansion) [Ye et al., 2013]. As such, the macrocell BS reduces the number of users it is required to serve, which improves the overall data rate for both the macrocell BS's users and also the user associated to the small-cell.

A popular low complexity approach for load balancing is known as biasing, where the user association is based on the biased received power. User association algorithms based on biasing have been proposed in [Ye et al., 2013] for the downlink, where it was surprisingly shown that user association based on biasing is near optimal, when the appropriate bias parameters are chosen. Alternative approaches based on various mixed integer programming problems have also been proposed in [Chen and Hu, 2012, Ye et al.,

2013, Andrews et al., 2013],

An important distinction between heterogeneous and traditional multiuser networks–affecting user association on both the uplink and downlink–is that users have approximately the same transmit power, while the BS transmit power can vary dramatically. This leads to a difference between the optimal user association for the uplink and downlink, which does not occur in traditional networks. To resolve this difference between the uplink and downlink user associations, Andrews [Andrews, 2013] has recently proposed that the uplink and the downlink be treated as different networks–a significant further departure from traditional macrocell-only design.

1.4 Cross-Layer Resource Allocation

Heterogeneous networks are interference-limited, which means that an important challenge is to efficiently allocate resources distributed over the network in order for BSs to support user demands. Important resources are frequency subcarriers, time slots, antennas in multiple-input multiple-output (MIMO) networks, and transmit power at each BS. There are two aspects to the resource allocation design challenge: allocation of a given BS's resources to associated users; and interference mitigation, for which algorithms are designed to provide each BS with the required resources. Importantly, resource allocation that ensures reliable and on-time transmission of data must consider not only each user's data rate but also the data network and queues servicing the wireless multiuser network. This leads to the concept of cross-layer resource allocation, where the standard networking layer separation does not hold.

Although cross-layer resource allocation has been proposed for earlier wireless multiuser networks, it has reached a new degree of importance in heterogeneous networks. This is due to the fact that the increased interference requires even more effective use of resources; in part to mitigate interference. In this section, we outline the effect of queueing and then survey state-of-the-art approaches to cross-layer user scheduling in wireless multiuser networks and, more generally, heterogeneous networks.

1.4.1 Queueing and the Data Network

While each user's achievable data rate is an important factor in the overall network performance, an equally important factor is the data demand. User data demand determines the arrival rate of packets to be transmitted via the wireless access network, and hence strongly affects the length of data queues and packet delay. In addition to the data rate of the wireless link, the packet arrival rate determines whether packet delay targets are met, which impacts on overall user satisfaction; for instance, failure to meet delay targets in wireless video transmission leads to buffering [Khalek et al., 2012].

Traditionally, wireless communication networks have been designed with a focus on information theoretic data rates. Unfortunately, this approach does not account for bursty data sources [Ephremides and Hajek, 1998]. To overcome the problem of bursty data sources, [Telatar and Gallager, 1995] derived the tradeoff between delay and channel capacity for Gaussian multiaccess channels. Later in [Berry and Gallager, 2002], the results in [Telatar and Gallager, 1995] were extended to fading channels albeit for point-to-point links. Throughput optimal rate allocation under a queue stability constraint was developed in [Yeh and Cohen, 2003], which lead to the "longest queue highest possible rate" policy. Finite buffer channels were considered in [Diggavi and Grossglauser, 2006], where the capacity of the compound channel consisting of a finite buffer queue and a discrete memoryless channel was obtained.

To characterize the practical performance of the data queue in heterogeneous networks, there are four key performance metrics: average queue length; average packet delay; the packet loss probability (PLP) due to buffer overflow; and queue stability. An important connection between the average queue length, average packet delay, and the probability of packet loss is Little's law [Zalesky, 2009], which states

$$L = \lambda W, \tag{1.1}$$

where L is the average number of packets in the queue, W is the average delay, and λ is the effective arrival rate, which accounts for both the user demand and also the packet loss probability.

Queue stability can be formalized in several ways and can be understood intuitively as ensuring that the queue length is bounded. In the case that the queue arrival and service processes are Markovian, then queue stability can be interpreted as positive recurrence [Norris, 1997]. More generally, when the arrival process is ergodic and the service process rate is bounded, then Neely et al. [Neely et al., 2002] proposed the notion of an overflow function

$$g(M) = \lim_{t\to\infty} \sup \frac{1}{t} \int_0^t \mathbf{1}_{[U(\tau)>M]} d\tau, \tag{1.2}$$

where $U(t)$ is the number of unprocessed packets at time t. The queue is then stable if $g(M) \to 0$ as $M \to \infty$.

1.4.2 Optimal Cross-Layer Resource Allocation

Driven by the need to account for both the behavior of physical layer channels and queueing dynamics, resource allocation in heterogeneous networks is a cross-layer design problem. Ideally, the cross-layer design problem is solved optimally via low-complexity distributed algorithms. Surprisingly, this ideal scenario exists in many practical scenarios.

In [Chiang et al., 2007a], Chiang et al. proposed a general framework for optimally solving cross-layer design problems using tools from convex optimization theory [Boyd and Vandenberghe, 2004]. In particular, they showed that a large number of cross-layer design problems can be written as a network utility optimization problem–commonly with the data rate as the utility function–with a decomposable structure [Palomar and Chiang, 2006]. This means that it is possible to solve the problems in a distributed fashion with no loss of optimality. In fact, solving the cross-layer design problem via decomposition has been shown to be applicable to cross-layer problems [Chiang et al., 2007a] involving routing, scheduling, power allocation and congestion control.

Unfortunately, obtaining the globally optimal solution by applying the decomposition technique in [Chiang et al., 2007a] requires that the utility function and constraints have an underlying separable convex structure. While this is the case for many problems, it is not true for all. In the case of non-separable or non-convex structure, the globally op-

timal solution cannot be efficiently found using distributed algorithms. In particular, the problem must be relaxed, which means that only locally optimal solutions can typically be found; often without bounds on the suboptimality. An important instance where this occurs is for data rate maximization subject to queueing delay constraints in heterogeneous networks [Chiang et al., 2007b]. Another pitfall of the network utility maximization techniques in low complexity implementations is that the numerical optimization algorithms are typically iterative; based on variations of Newton's method [Boyd and Vandenberghe, 2004].

The alternative to iterative numerical optimization algorithms is to exploit heuristic approaches [Fattah and Leung, 2002, Conte et al., 2010, Anton-Haro et al., 2006, Liu et al., 2006b, Carneiro et al., 2004]. While heuristics are not globally optimal solutions to the network utility maximization problem, they have the dual benefits of low complexity–the heuristic is not an iterative algorithm–and analytical tractability, which means that analytical performance guarantees can be derived.

1.4.3 Cross-Layer User Scheduling

User scheduling is an important resource allocation problem; in fact, user scheduling is the main factor in determining the achievable data rate of users in interference-limited heterogeneous networks [Andrews, 2013]. On the downlink, the user scheduling problem consists of selecting one (or more) users to be serviced by the associated BS in a given time slot. This is a cross-layer design problem as BSs are required to ensure that each user receives a target data rate, satisfies packet delay targets, and queue lengths are bounded. As in the case of general cross-layer design, there are two key approaches to cross-layer user scheduling: network utility maximization; and the heuristic approach.

In [Hou and Kumar, 2010], the network utility maximization approach was applied to the cross-layer user scheduling problem. In particular, Hou et al. considered time varying wireless links, where the BS services N users. The BS either fully meets the demand of each user, or does not service the user at all. Under the assumption of a finite state channel model, the network utility maximization problem was formulated such that the sum long-term utility of each user is maximized subject to the long-term service rate

target being met.

The network utility maximization for the scheduling problem in [Hou and Kumar, 2010] was then solved by developing an online scheduling policy, which asymptotically achieved the globally optimal solution of the network utility maximization problem. Unfortunately, there are two factors that affect the practical use of the online scheduling policy in low-complexity networks: the policy is based on an iterative numerical algorithm; and the network utility maximization is formulated on the basis of the long-term strictly concave utility and long-term service-rate, which does not provide any guarantees on the network behavior in the short-term and for general non-strictly concave utility functions.

In situations where the utility yields a non-convex network utility maximization problem or computational complexity is limited, the heuristic approach is an effective alternative. There are three key heuristic policies for the cross-layer user scheduling problem, where one user is selected in each time slot: opportunistic [Fattah and Leung, 2002], weighted data rate [Liu et al., 2006b]; and cumulative distribution function (CDF)-based [Park et al., 2005].

Opportunistic scheduling is the simplest scheduling policy, where the queue state is neglected and the user with the highest data rate is selected [Fattah and Leung, 2002]. To account for delay, the weighted data rate policy was proposed to select the user with the highest weighted data rate, where the weight is chosen to be a function of the delay. In [Liu et al., 2006b], the weight is determined by the total packet delay, which means that this policy is able to trade off the data rate for a lower total packet delay. Unfortunately, this policy requires knowledge of the total packet delay for each user's queue, which can have a wide variation and hence a long description. In [Park et al., 2005], the CDF-based scheduling policy was proposed, which offered a simple method to ensure that target user selection probabilities were achieved.

Cross-layer user scheduling has also been proposed for heterogeneous networks. An important application is in coordinated multipoint (CoMP) networks, where BSs cooperate to improve service to cell-edge users, which typically experience a high degree of interference. In [Seifi et al., 2011], user scheduling policies were developed. Resource allocation for CoMP joint transmission in OFDMA networks was developed in [Choi et al.,

2011], while weighted sum-rate maximization with discrete power control was proposed in [Zhang et al., 2011].

1.5 Interference Mitigation

Without interference mitigation, each BS in the heterogeneous network does not have the resources–in the form of spectrum and time slots–required to effectively service its users. Interference mitigation strategies have three dimensions to exploit: frequency; time; and space. As such, heterogeneous network interference mitigation strategies are based on frequency reuse (frequency dimension), BS scheduling (time dimension), and network MIMO (space dimension). In this section, interference mitigation strategies exploiting frequency reuse, BS scheduling, and network MIMO are surveyed with a focus on approaches applicable to multiuser and heterogenous networks.

1.5.1 Frequency Reuse

In cellular networks using OFDMA, nearby BSs are allocated different frequency bands to reduce the interference between adjacent cells. This approach can significantly improve the performance of cell-edge users. Traditionally, there are two common approaches [Novlan et al., 2011]: fractional frequency reuse (FFR), where users in the cell-interior are allocated a common frequency and the bandwidth for cell-edge users is partitioned over the network; and soft frequency reuse (SFR), where the bandwidth for cell-edge users is partitioned the same as for FFR, however the interior users may share frequency bands with cell-edge users in nearby cells.

With the addition of small-cells in heterogeneous networks, more sophisticated frequency reuse strategies are required. Two common approaches are multicarrier deployment and carrier aggregation [Damnjanovic et al., 2011]. In multicarrier deployment, small-cells are allocated different frequency bands to the macrocell, which completely removes interference between tiers. On the other hand, carrier aggregation provides the macrocell with its own bandwidth and shares the remainder with the small-cells, which improves frequency reuse.

While multicarrier deployment and carrier aggregation are simple to implement, near-optimal frequency reuse strategies can achieve significantly higher system capacity. In [Saquib et al., 20], a static frequency reuse scheme was proposed, which partitions each cell into 12 zones. In [Liang and Goldsmith, 2007], an adaptive frequency reuse scheme was propose based on the channel strength. Han et al. [Han et al., 2009] proposed a joint optimization of transmit power, frequency, channel and deployment location for each BS.

1.5.2 Base Station Scheduling

Particularly if SFR is employed, interference for cell-edge users can be severe. For these users, it is highly desirable for nearby BSs to cooperate by scheduling transmissions in orthogonal time slots. A key technology for base station scheduling on the downlink is known as the coordinated scheduling category of coordinated multipoint (CoMP)– currently standardized in LTE [Irmer et al., 2011].

Due to the fundamental similarities between user scheduling and CoMP scheduling, most scheduling policies have the same structure in the common scenario of single antenna BSs and users. In particular, the schedulers in [Fattah and Leung, 2002, Liu et al., 2006b, Park et al., 2005, Hou and Kumar, 2010] can be directly applied to the CoMP scheduling problem. Importantly, sufficient backhaul must be available in order to ensure that the centralized computations can be performed.

1.5.3 Network MIMO

The third dimension for interference mitigation is space, where BSs, relays and remote radio heads cooperate to form a distributed MIMO network, known as network MIMO. There are a number of ways that MIMO can be exploited in heterogeneous networks; in particular, four common scenarios are:

1. relays and remote radio heads, used to improve coverage;
2. multiuser MIMO, for the single-cell;
3. coordinated multipoint (CoMP), for cell-edge users in multi-cell environments;

4. multi-cell BS cooperation.

Relays have been proposed to improve coverage and reliability of the network. Early work exploited information theoretic techniques to evaluate the performance. In particular, [Cover and El Gamal, 1979] derived the achievable rate based on the estimate-and-forward relaying protocol. In [Laneman et al., 2004, Sendonaris et al., 2003] a number of practical half-duplex relaying protocols were proposed and analyzed, known as amplify-and-forward and decode-and-forward. More recently, relaying schemes have been proposed for networks with multiple relays [Gastpar et al., 2002] and also relays with multiple antennas [Bolcskei et al., 2006].

Since the seminal work of Foschini [Foschini and Gans, 1998] and Telatar [Telatar, 1999], MIMO techniques for multiuser networks have become ubiquitous. In [Caire and Shamai, 2003], Caire and Shamai introduced multiuser MIMO, where a BS with multiple antennas transmitted data to several users simultaneously. In fact, for users with a single antenna the achievable sum-rate for multiuser MIMO scales linearly with the minimum of the number of users and number of transmit antennas. Multiuser MIMO has since been standardized in LTE and LTE-A [Motorola, 2007], and is the basis for many multi-cell BS cooperation techniques [Gesbert et al., 2010].

The basic mathematical model for multiuser MIMO with n single antenna users and BS with m transmit antennas is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1.3}$$

where $\mathbf{y} \in \mathbb{C}^n$ is the receive vector with elements corresponding to each user, $\mathbf{H} \in \mathbb{C}^{n \times m}$ is the channel matrix, $\mathbf{x} \in \mathbb{C}^m$ is the transmitted data vector, and $\mathbf{n}$ is the noise. In order to mitigate inter-user interference, the desired data vector $\mathbf{u}$ is transformed to obtain the transmitted data vector $\mathbf{x}$. The optimal transformation is an iterative procedure known as dirty paper coding [Caire and Shamai, 2003]. In practice linear precoding is used, with $\mathbf{x} = \mathbf{V}\mathbf{u}$ where $\mathbf{V}$ is a linear transformation. Popular choices of $\mathbf{V}$ are the Moore-Penrose pseudoinverse (as in zero-forcing precoding [Caire and Shamai, 2003]), and regularized channel inversion [Peel et al., 2005], which is closely related to the minimum mean square

error receiver. An alternative high performance sub-optimal transformation with higher complexity is vector-perturbation, which was proposed in [Hochwald et al., 2005].

Multi-cell BS cooperation is the most general form of network MIMO, where BSs cooperate to form large virtual MIMO networks to service users within multiple cells simultaneously. A significant feature of multi-cell BS cooperation is that the cell-boundaries in fact disappear within the coordinating cells–there is no inter-cell interference.

The link between the downlink of a multi-user MIMO network and multi-cell BS cooperation scenarios was first developed in [Shamai and Zaidel, 2001] (see [Gesbert et al., 2010] for a detailed historical account), where it was observed that the two networks are equivalent if individual BS power constraints were neglected. Since this initial work, developments in resource allocation for multi-user MIMO with per-antenna power constraints [Wiesel et al., 2008] has further connected the two scenarios.

Recently, practical schemes for multi-cell BS cooperation have been pursued. A key desiderata for cooperation schemes is that they are distributed, which is achieved using two main techniques: optimization theory; and game theory. Iterative optimization schemes have been proposed in [Gesbert et al., 2007, Li and Liu, 2006]. In [Han et al., 2007] non-cooperative game theory was exploited to yield distributed resource allocation policies.

The major difficulty in using network MIMO techniques is that channel state information (CSI) is required at the BSs to mitigate inter-user interference on the downlink. In common frequency division duplex (FDD) networks, the uplink and the downlink use different frequency bands. As such, the CSI must be *fed back* from the users to the BSs. In the next section, we focus on state-of-the-art approaches for backhaul and CSI feedback in network MIMO.

1.6 Providing Side Information

To provide the capability for BS cooperation, side information is required in the form of other users' encoded data and channel state information. In particular, BSs need to be connected to each other via backhaul, and to users–potentially serviced by other BSs–via

feedback links. While feedback techniques have been developed for some time–often for the application of MIMO–new approaches are required in heterogeneous networks, where network performance is limited by the sheer volume of required side information. Importantly, similar techniques are also applicable for storage in cloud networks [Papailiopoulos et al., 2012]. In this section, we outline the current approaches and challenges in providing backhaul and feedback in heterogeneous networks.

1.6.1 Backhaul

In multi-cell cooperation schemes, a key element is the capability for BSs to share channel state information and data symbols in order to perform coordinated scheduling, beamforming or exploit network MIMO techniques. Sharing of information between BSs is achieved by backhaul in the form of either fibre or wireless links.

The requirement of backhaul presents many new challenges including who provides the backhaul, what information to share between BSs, and ultimately the cluster size of the cooperating BSs. In the case of femtocells, ISPs are envisaged to provide backhaul [Andrews et al., 2012]; however, this will force the ISPs to carry additional traffic. More generally, for in-building applications ethernet can support the backhaul requirements. On the other hand, for outdoor deployment of small-cells, microwave and millimeter wave radios have been proposed [Hoadley and Maveddat, 2012].

1.6.2 Feedback

While backhaul provides the capability for BSs to share information, feedback provides the capability for users to share information with BSs, which is crucial in frequency division duplex (FDD) networks exploiting network MIMO. The basic problem in limited feedback design is to determine the feedback rate required in order to achieve a given data rate. Early work focused on feedback of scalar channel state information corresponding to single antenna links, while more recently focus has been on feedback of vector channel state information in multi-cell MIMO networks.

The main approach for limited feedback design is to quantize the channel state infor-

mation (either scalar or vector). In order to perform quantization, both a quantization codebook and a quantization rule are required. In the case of MIMO links, the channel state information is a vector, which means that vector quantization is required. For single user MIMO channels exploiting beamforming, the quantization rule is [Love et al., 2003]

$$\hat{h} = \arg\min_{\hat{h}\in\mathcal{F}} |\mathbf{h}^{\dagger}\hat{\mathbf{h}}|^2, \tag{1.4}$$

where $\mathbf{h}$ is the normalized channel (i.e. $\|\mathbf{h}\| = 1$). In [Love et al., 2003], the Grassmannian criterion was proposed

$$\min\max_{i\neq j} |\hat{\mathbf{h}}_i^{\dagger}\hat{\mathbf{h}}_j|^2, \tag{1.5}$$

where $\hat{\mathbf{h}}_i, \hat{\mathbf{h}}_j \in \mathcal{F}$. In general, finding codebooks that optimize (1.5) is a hard problem as it amounts to constructing Grassmannian frames, which are notoriously difficult to find. Structured approaches to constructing quantization codebooks have many subtle connections with algebra, combinatorics, and geometry, which are elaborated further in [Love et al., 2003, Strohmer and Heath, 2003, Xia et al., 2005].

A special class of well-studied Grassmannian frames is known as equiangular tight frames (ETFs) [Strohmer and Heath, 2003]. In this case

$$|\hat{\mathbf{h}}_i^{\dagger}\hat{\mathbf{h}}_j|^2 = \frac{N - N_t}{N_t(N-1)}, \tag{1.6}$$

where N is the number of codewords and N_t is the number of transmit antennas. In fact, codebooks that are ETFs minimize (1.5). Unfortunately, these codebooks only exist when the number of codewords satisfy $N \leq N_t^2$ [Strohmer and Heath, 2003]. Explicit constructions have yielded a large number of special cases ranging from difference set-based techniques [Xia et al., 2005], to harmonic frames [Strohmer and Heath, 2003]; although most constructions rely on deep links to combinatorial design theory [Colbourn and Dinitz, 2007]. In [Tropp et al., 2005], ETFs were constructed using numerical search based on alternating minimization. The existence of ETFs for all N and N_t such that $N \leq N_t^2$ remains an open question.

Alternative structured approaches for quantization codebook design have also been proposed, which are based on the Fourier codebook [Love et al., 2003], Kerdock codes [Inoue and Heath, 2009], lattices [Ryan et al., 2009] and the modified Lloyd algorithm [Xia et al., 2005]. While suboptimal and impractical, random approaches have also been proposed and analyzed in [Mukkavilli et al., 2003].

For multiuser MIMO, obtaining quantization codebooks has been even more challenging than for single user MIMO. In [Heath et al., 2009] it was shown that the Grassmannian criterion is still applicable in the case that zero-forcing precoding is employed. The key difference between single and multiuser MIMO quantization codebook design is that inter-user interference is present. As such, significantly larger codebooks are required. Random codebooks were proposed and analyzed in [Jindal, 2006], and a rate-distortion analysis approach for random codebooks was performed in [Ding et al., 2007].

To reduce the search complexity and storage requirements, structured multiuser MIMO quantization codebooks have also been proposed. In particular, Kerdock codes [Inoue and Heath, 2009], progressive refinement techniques [Heath et al., 2009], and Lloyd algorithm-based approaches have been considered. An alternative approach where a precoder is selected from a finite set of choices instead of quantizing the channel at each user has also been proposed [Love and Heath, 2005]. The quantization codebook employed in LTE is described in [Heath et al., 2009].

In multi-cell MIMO networks, the feedback problem is even more complex as different levels of feedback may be provided to different BSs. To this end, adaptive feedback allocation schemes have been proposed and analyzed in [Bhagavatula and Heath, 2011, Lee and Shin, 2011, Hou and Yang, 2011].

1.7 Physical-Layer Security

Although the data rate is an important performance metric, in many applications it is not the only one that determines the effectiveness of the network; one example is cross-layer design. Another case where data rate is not the only criterion is in network security. Secure transmission of data is of practical necessity in not only military applications, but

also in the protection of individual privacy [Shiu et al., 2011]. The traditional approach to the security problem is to employ cryptographic techniques that exploit the computational deciphering of the codes. Unfortunately, with the computational advances following Moore's law and the vulnerability of the cryptographic algorithms to future theoretical advances it is highly desirable to ensure perfect secrecy.

Perfect secrecy was first introduced by Shannon, and is achieved by ensuring that the mutual information of the received message at the eavesdropper is zero [Mukherjee et al., 2013]. Effectively, this means that the eavesdropper gains no additional information from the received signal, which is used to decode the message. Shannon's result was then extended by Wyner to the wiretap channel, which is based on the degraded broadcast channel.

More recently, fading channels and MIMO techniques have been considered. In particular, the secrecy capacity was derived for MIMO channels with eavesdropper–known as the MIMO wiretap channel–in [Khisti and Wornell, 2010a, Khisti and Wornell, 2010b]. An important technique for the MIMO wiretap channel is artificial noise transmission [Goel and Negi, 2008], where noise is transmitted into subspaces that are orthogonal to the signal for the intended user. As such, only the eavesdropper experiences a degradation in channel quality.

Physical layer security techniques have also been extended to heterogeneous networks. In particular, secrecy capacity scaling laws were derived in [Koyluoglu et al., 2012, Liang et al., 2009, Vasudevan et al., 2010]. In [Zhou et al., 2011], the throughput cost was analyzed.

1.8 Focus of the Thesis and Overview of Contributions

This thesis focuses on three areas within the context of Sections 1-7 that will impact future wireless multiuser and heterogeneous networks. The first area is cross-layer scheduling, where low complexity approaches are considered for both single-cell and CoMP networks that incorporate reliable physical layer transmission and finite buffer queues. The second area is limited feedback quantization codebook design in both multiuser MIMO

and CoMP, with a focus on structured codebooks with low storage, and low construction and search complexity. The final area is resource allocation for physical layer security multiuser MIMO, where the algorithms are proposed and analyzed in terms of the trade-off between achieving a high sum-rate and achieving secrecy.

In each area, algorithms are developed to improve performance and efficiently use network resources. The major contributions in this thesis are the following algorithms.

1. Low complexity cross-layer scheduling policies for single-cell multiuser networks aided by a relay and CoMP.

2. Structured quantization codebook design in multiuser MIMO and CoMP with limited feedback.

3. Precoder design and power allocation to maximize the secrecy sum-rate in multiuser MIMO.

In each area, the proposed algorithms are detailed and analyzed. In particular, optimal or near-optimal solutions are obtained via low complexity techniques, which are applicable in a range of multiuser networks. Each aspect contributes to reliable and high data-rate transmission in large-scale wireless multiuser and heterogeneous networks. The algorithms proposed in this thesis improve the practicality or performance in systems with a range of physical and computational resources.

1.8.1 Overview of Major Contributions

In this section, the structure of this thesis is outlined with a detailed overview of the key contributions.

Chapter 2: Cross-layer User Scheduling in Single-cell Relay Networks

In this chapter, a cross-layer user scheduler is proposed in multiuser networks aided by a relay. In order to guarantee performance targets are met, analytical expressions are derived for delay, symbol error probability, and packet loss due to buffer overflow. In particular, the key contributions are:

1. *Section 2.3* develops the notion of the delay in packet scheduling (DPS), and gives the proposed user scheduler.

2. *Section 2.4.2* gives the expression for the probability mass function of the DPS.

3. *Theorem 2.1* gives the cumulative density function of the end-to-end signal-to-noise ratio (through the relay network) for the scheduled user.

4. *Equation (2.19)* gives the symbol error probability of the scheduled user, which is applicable to several different modulation formats.

5. *Theorem 2.2* gives the packet loss probability due to buffer overflow for each user in networks with finite data buffers.

Chapter 3: Cross-layer Scheduling in Coordinated Multipoint

In this chapter, a cross-layer base station scheduler is proposed for the downlink of a coordinated multipoint network with finite buffers. Performance guarantees are derived for the packet loss probability due to buffer overflow. To efficiently describe the tradeoff between the packet loss probability for each base station and the packet arrival rate, the concept of the τ-achievable region is developed. The key contributions of this chapter are detailed as follows:

1. *Section 3.3.1* gives the proposed cross-layer base station scheduler.

2. *Proposition 3.1* gives the bound for the packet loss probability due to buffer overflow for each user.

3. *Definition 3.1* gives the concept of the τ-achievable region.

4. *Theorem 3.1* gives the τ-achievable region for the proposed scheduler.

5. *Section 3.4* gives the proposed network design strategy to select both the parameters for the scheduling policy and the transmission time of each slot.

Chapter 4: Structured Limited Feedback Codebook Design for Multi-user MIMO

In this chapter, the problem of constructing different quantization codebooks at each user in multiuser MIMO is considered. This ensures that rank deficiency of the quantized

channel matrix is avoided, which significantly improves performance when zero-forcing precoding is employed. Two structured constructions for the codebooks are proposed: the first construction is based on the Householder transform; and the second construction is based on group representation theory. In particular, the key contributions in the chapter are:

1. *Table 4.1* gives the first codebook construction method based on the Householder transform.

2. *Table 4.2* gives the second codebook construction method based on group representation theory.

3. *Theorem 4.1* gives conditions for the uniqueness of the transformed codebooks at each user.

4. *Proposition 4.2* gives new sparsity properties for the Fourier codebook after applying the Householder-based construction in *Table 4.1*.

5. *Equation (4.22)* gives the rate-outage bound when a common codebook is employed to provide a baseline comparison with the proposed constructions of different codebooks.

6. *Theorem 4.2* gives new properties of optimal base codebooks in terms of the rate-outage bound.

Chapter 5: Structured Limited Feedback Codebook Design for Coordinated Multipoint

In this chapter, the problem of constructing variable dimension quantization codebooks for multiuser MIMO is considered. First, a low complexity codebook construction is proposed, which has linear complexity in both codebook size and dimension Next, optimal codebooks based on cyclic difference sets are used to construct codebooks with low storage requirements. The key contributions in this chapter are:

1. *Section 5.4* gives the proposed low complexity construction with linear complexity in both size and dimension.

2. *Section 5.5* gives the proposed construction with low storage requirements based on the theory equiangular tight frames.

Chapter 6: Resource Allocation in Multiuser MIMO for Physical Layer Security

In this chapter, resource allocation algorithms are proposed to achieve physical layer security in multiuser MIMO networks. In particular, joint precoder design and power allocation algorithms are developed. The key contributions in this chapter are:

1. *Theorem 6.1* gives a general expression for the achievable secrecy sum-rate in multiuser MIMO with linear precoding.

2. *Theorem 6.2* gives the optimal secrecy sum-rate achieved by regularized channel inversion precoding in terms of the signal-to-noise ratio, for the large-system regime.

3. *Table 6.1* gives algorithms for power allocation, and joint power allocation-precoder design.

1.9 Publications

This section is a list of publications based on work that was done during the period of PhD candidature. Importantly, there is no overlap in work between any of the journal papers and conference proceedings, with the exception of

1. conference paper (C4) that forms part of Sections III, IV, V, and VII in journal paper (J4);

2. and Section IV in conference paper (C5) that overlaps with part of Section IV in journal paper (J3).

1.9.1 Journal Papers

(J1) M. Egan, I.B. Collings, "Bandwidth-Limited Distributed Estimation of Level Crossings", in preparation.

(J2) M. Egan, G. Peters, I. Nevat, I.B. Collings, "On the Feasibility of Third-Party Operated Micro-Networks", in preparation.

(J3) M. Egan, C.K. Sung and I.B. Collings, "Structured and Sparse Limited Feedback Codebooks for Multiuser MIMO", IEEE Trans. on Wireless Communications, Vol. 12, No. 8, pp. 3710-3721, August 2013.

(J4) M. Egan, P.L. Yeoh, M. Elkashlan and I.B. Collings, "A New Cross-Layer User Scheduler for Wireless Multimedia Relay Networks", IEEE Trans. on Wireless Communications, Vol. 12, No. 1, pp. 301-311, January 2013.

(J5) G. Geraci, M. Egan, J. Yuan, A. Razi and I.B. Collings, "Secrecy Sum-Rates for Multi-User MIMO Regularized Channel Inversion Precoding", IEEE Trans. on Communications, Vol. 60, No. 11, pp. 3472-3482, November 2012.

1.9.2 Conference Proceedings

(C1) M. Egan and I.B. Collings, "Base Station Cooperation for Queue Stability in Wireless Heterogeneous Cellular Networks", in the Proc. of the IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), London, UK, September 2013.

(C2) M. Egan and I.B. Collings, "Low Complexity Quantization Codebooks for CoMP", in the Proc. of the IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), London, UK, September 2013.

(C3) M. Egan, P.L. Yeoh, M. Elkashlan and I.B. Collings, "A Coordinated Multipoint Scheduler for Packet Loss Reduction", in the Proc. of the IEEE Int. Vehicular Technology Conf. (VTC) - Spring, Dresden, Germany, June 2013.

(C4) M. Egan, P.L. Yeoh, M. Elkashlan and I.B. Collings, "A New Cross-Layer User Scheduler for Delay and Symbol Error Probability in Wireless Multimedia Relay Networks", in the Proc. of the 14th Australian Communications Theory Workshop (AusCTW), Adelaide, Australia, February 2013.

(C5) M. Egan, C.K. Sung and I.B. Collings, "Codebook Design for the Finite Rate MIMO Broadcast Channel with Zero-Forcing Precoding", in the Proc. of the IEEE Int. Conf. on Global Communications (GLOBECOM), Wireless Communications Symposium, Houston, USA, December 2011.

(C6) M. Egan, I.B. Collings, W. Ni and C.K. Sung, "User Scheduling for the Broadcast Channel Using a Sum-Rate Threshold", in the Proc. of the IEEE Int. Conf. on Communications (ICC), Wireless Communications Symposium, Kyoto, Japan, June 2011.

Chapter 2

Cross-layer User Scheduling in Single-cell Multiuser Wireless Relay Networks

This chapter proposes a new user scheduler for multiuser networks that are aided by a relay. The scheduler is designed to account for delay, symbol error probability (SEP), and packet loss probability (PLP) due to buffer overflow. A cross-layer scheduling approach is developed for the downlink to balance these system metrics.

2.1 Introduction

THE explosive demand for real-time audio and video streaming drives the need for new high data-rate transmission strategies in multiuser wireless networks supporting multimedia applications [Akyildiz et al., 2007, Fattah and Leung, 2002]. The defining property of these networks is low delay tolerance, in addition to throughput constraints [Akyildiz et al., 2007]. Practical implementations of wireless multimedia networks are currently under consideration in current and emerging standards such as the IEEE 802.16 [Liu et al., 2006b, So-In et al., 2009, So-In et al., 2010]. Wireless multimedia relay networks (WMRNs) are one important class of such networks [Johansson and Xiao, 2006, Ng and Yu, 2007, Tang and Zhang, 2007]. In this chapter, the important problem of scheduling in WMRNs with multiple data classes and multiple users is addressed. Not only does it call for a cross-layer approach to address the multimedia data, but it must also have a cross-user aspect that takes into account the instantaneous channel states of all the users.

Cross-layer scheduling has been successfully applied to other wireless multimedia networks. There are two main approaches: low-complexity heuristic policies [Liu et al., 2006a, Kobayashi and Caire, 2006, Anton-Haro et al., 2006, Eryilmaz and Srikant, 2004, Park et al., 2005], and network utility maximization (NUM) based algorithms [Hou and Kumar, 2010, O'Neill et al., 2008]. A common example of the heuristic approach is the weighted sum-rate scheduler, where the weight is dependent on a function of total packet delay [Liu et al., 2006a] or the number of packets in the queue [Kobayashi and Caire, 2006]. In contrast, NUM maximizes the average utility subject to a long-term minimum service rate for each user [Hou and Kumar, 2010, O'Neill et al., 2008]. While NUM achieves long-term utility optimality, it requires the repeated solution of a convex optimization problem using iterative sub-gradient methods. This may not be practical in low-complexity systems.

In this chapter, a new cross-layer scheduler is proposed for the WMRN that falls within the heuristic approaches. The new scheduler is specifically designed for the multiuser situation. Unlike the single user case, the overall delay is not only affected by the length of the queue, but can in fact be dominated by the channel state for each user, relative to all the others. To address this, a new scheduler is proposed based on a weighted function of the signal-to-noise ratio (SNR) for each user, and the delay in packet scheduling (DPS) for each user. The DPS for a user is defined as the number of potential scheduling opportunities that have elapsed for the packet at the head of the user's queue.

To demonstrate the performance of the proposed scheduler, it is applied to the popular WMRN based on cooperative dual-hop amplify-and-forward (AaF) relaying. The majority of work in the literature related to scheduling in these WMRNs consider opportunistic equal weight scheduling, where the user with the highest end-to-end SNR is scheduled (see, for example, [Yang et al., 2011, Kim et al., 2011, Ding et al., 2011] and the citations therein). Of particular interest is [Yang et al., 2011] where the outage probability and the symbol error probability (SEP) of opportunistic equal weight scheduling were derived. For the proposed scheduler, the fact that the DPS evolves according to a Markov chain is observed, and then used to calculate both the PMF of the DPS and the SEP. The buffer state (number of packets) and the packet loss probability due to overflow

are also calculated. These quantities are required to calculate the average total delay of each packet and the throughput.

The network is analyzed under the assumption of independent and identically distributed Rayleigh fading and homogeneous Poisson arrivals. First, the probability mass function (PMF) of the DPS is derived by constructing the augmented transition matrix of a truncated multidimensional Markov chain. Then, an accurate approximation is derived for the average SEP of the scheduled user with uncoded symbols. The analysis holds for an arbitrary number of users K and any scheduling policy where the weights are a function of the DPS for each user. Simulations are then performed that confirm the analysis. The proposed scheduler is shown to achieve up to a 30% improvement in the probability the DPS satisfies a given constraint, with negligible degradation in the SEP compared to the opportunistic equal weight scheduler in [Yang et al., 2011].

To further characterize the performance of the scheduler, a new analytical expression is derived for the PLP due to buffer overflow. Currently, this has not been addressed in the multiuser scheduling scenario under consideration. Simulation demonstrate that the proposed scheme can outperform the equal weight scheduler [Yang et al., 2011] by 85%. The analysis and simulations show that that the PLP does not always improve with increasing buffer size. This is shown to be avoided by decreasing the transmission time.

Finally, the throughput is simulated when a convolutional code is used. The throughput is dependent on the SEP of the scheduled user, and the average PLP. When the transmission time is sufficiently small, the SEP dominates the throughput. In contrast, when the transmission time is sufficiently large, the PLP dominates. Simulations show that the throughput can be significantly improved by optimizing the transmission time and the code rate. In particular, the throughput can be increased by over 20% when the transmission time is reduced from 4 ms to 2 ms, at a code rate of $1/4$.

2.2 System Model

Consider the WMRN (illustrated in Fig. 2.1) where a base station (BS) with K first-in first-out (FIFO) data queues transmits to K corresponding users with the aid of an AaF[1] relay.

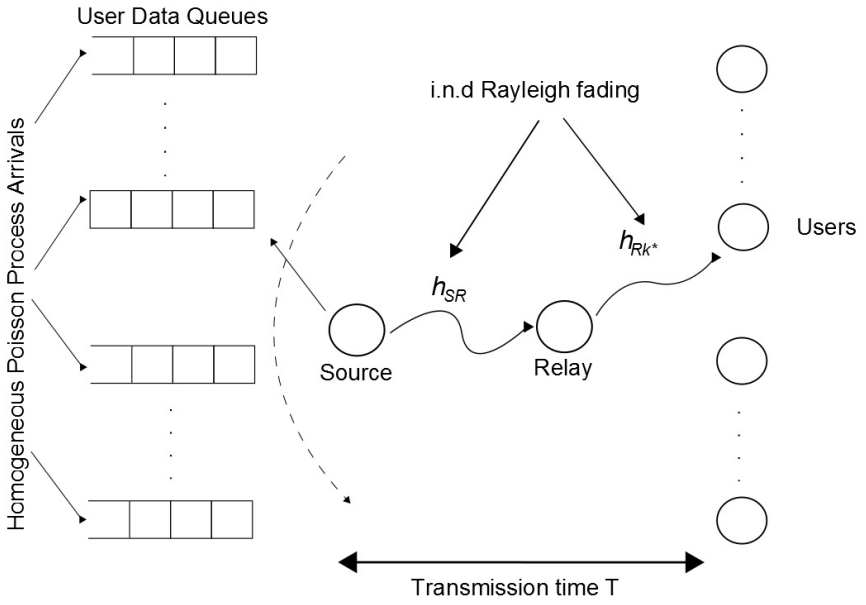


Figure 2.1: WMRN system model.

In the proposed cross-layer scheduling policy, a single user with the largest weighted SNR is scheduled for transmission in each scheduling opportunity. Independent non-identically distributed (i.n.d) block Rayleigh fading is assumed in the two-hop relay links with a coherence time of T_c seconds.

2.2.1 MAC Layer Architecture

The BS has K finite queues with buffer size B, each corresponding to a distinct user. A user's packet is lost if the buffer for the queue is full and a new packet arrives. The arrival process of the packets for each queue is assumed to be a homogeneous Poisson process with rate λ_k, $k = 1, \ldots, K$, where each k corresponds to a different queue. The probability that n packets arrive in an interval of time T for the k-th user is then given by

$$\Pr(N_k(T) = n) = \frac{e^{-\lambda_k T}(\lambda_k T)^n}{n!}. \tag{2.1}$$

[1] AaF is a practical relaying protocol due to its ease of implementation and low power consumption since there is no need for digital decoding at the relay [Laneman et al., 2004]. This makes it preferable over decode-and-forward in low-complexity implementations where digital hardware is not available.

Packets can be re-requested with the caveat that the arrival of the re-requested packet is consistent with the Poisson arrival process. The transmission time T is the same for all users.

Prior to Section 2.5, backlogged queues are assumed such that at least one packet is always available. As a result, the BS is never silent. This assumption is also made in [Sharif and Hassibi, 2007]. This restriction in Section 2.6 where the PMF of the buffer state and the PLP due to buffer overflow are derived.

2.2.2 Physical Layer Architecture

The BS and the relay each transmit for $T/2$ seconds in half duplex mode such that the total transmission time from the BS to the scheduled user is T seconds, where $T \leq T_c$. The transmission time is chosen such that the BS has knowledge of both the BS-relay and relay-user links for scheduling purposes. In the BS-relay link, the received signal at the relay is given by

$$y_R = \sqrt{E_S} h_{SR} x + z_R, \tag{2.2}$$

where E_S is the transmit power at the source, h_{SR} is the Rayleigh fading channel coefficient between the source and the relay, x is the transmitted symbol using binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK) or M-ary pulse amplitude modulation (M-PAM), and z_R is the additive white Gaussian noise (AWGN) with one-sided power spectral density N_0. In the relay-to-user link, the received signal at the scheduled user, denoted by $k^* \in \{1, \ldots, K\}$, is given by

$$y_{k^*} = \sqrt{E_R} h_{Rk^*} \beta y_R + z_{k^*}, \tag{2.3}$$

where E_R is the transmit power at the relay, h_{Rk^*} is the Rayleigh fading channel between the relay and the scheduled user, z_{k^*} is the AWGN with one-sided power spectral density

N_0, and β is the relay amplification factor defined as

$$\beta = \sqrt{\frac{1}{E_{SR}|h_{SR}|^2 + cN_0}}. \tag{2.4}$$

In (2.4), set $c = 1$ for the case where noise power is included in the relay amplification factor and set $c = 0$ for the case where the noise power is ignored [Yang et al., 2011].

The end-to-end SNR of the scheduled user is written as

$$\gamma_{eq} = \frac{\gamma_{SR}\gamma_{Rk^*}}{\gamma_{SR} + \gamma_{Rk^*} + c} \tag{2.5}$$

where γ_{SR} is the instantaneous SNR in the source-to-relay link and γ_{Rk^*} is the instantaneous SNR in the relay-to-user link. The effect of path loss is incorporated into the instantaneous SNRs such that $\gamma_{SR} = d_S^{-\eta} E_S |h_{SR}|^2 / N_0$ and $\gamma_{Rk^*} = d_R^{-\eta} E_R |h_{Rk^*}|^2 / N_0$, where d_S is the distance between the source and the relay, d_R is the distance between the relay and the scheduled user, and η is the path loss exponent. As both the BS-relay and relay-to-user links experience i.n.d. Rayleigh fading, the probability density functions (PDFs) of the instantaneous SNRs are written as

$$f_{\gamma_{SR}}(\gamma) = \frac{1}{\overline{\gamma}_1} e^{-\frac{\gamma}{\overline{\gamma}_1}}, \tag{2.6}$$

$$f_{\gamma_{Rk}}(\gamma) = \frac{1}{\overline{\gamma}_2} e^{-\frac{\gamma}{\overline{\gamma}_2}},\ k = 1, 2, \ldots, K. \tag{2.7}$$

where $\overline{\gamma}_1 = E[\gamma_{SR}]$ is the average SNR in the source-to-relay link and $\overline{\gamma}_2 = E[\gamma_{Rk}]$, $\forall k \in \{1, \ldots, K\}$, is the average SNR in the relay-to-user links, with $E[\cdot]$ denoting the expectation.

2.3 Proposed Cross-Layer Scheduling Policy

In this section, the new user scheduling policy is detailed. The scheduling policy selects the user with the largest weighted SNR of the second hop. The weight is a function of the DPS. The DPS is formally defined as follows.

Definition 2.1. *The DPS of user k is the number of potential scheduling opportunities that have elapsed for the packet at the head of user k's queue.*

Note that a packet can only be scheduled at the front of a user's queue. As a result, only delays of the packets at the front of each user's queue are required for the new scheduler's computations. The header size of each packet can then be significantly reduced in long queues compared with the scheme in [Liu et al., 2006b], as time stamps with a small number of bits are sufficient. The reduction is due to the impact of the large variation in total packet delay on the scheme in [Liu et al., 2006b], caused by the dependence on the number of packets in the queue when the packet arrives.

While the approach is heuristic, it is practical and efficient. Since the modulation scheme is fixed, the instantaneous SEP only varies when the instantaneous SNR varies. As a result, only the instantaneous SNR is required in the scheduling policy to account for the SEP requirements.

The proposed scheduling policy is given by

$$k^* = \arg \max_{k=1,\ldots,K} \gamma_{Rk} W_k, \tag{2.8}$$

where

(i) k^* is the scheduled user;

(ii) γ_{Rk} are the instantaneous SNRs of the relay-to-user links, where $k = 1, \ldots, K$;

(iii) W_k is the weight satisfying

$$W_k(s_k) \geq 0,\ k = 1, \ldots, K \tag{2.9}$$

where $W_k(s_k)$ is an arbitrary function of s_k that is the number of potential scheduling opportunities that have elapsed for the k-th user's packet. The BS-relay link is not explicitly present in the scheduling policy as it is the same for all users.

Users in multimedia applications are often characterized as real time (RT) or best effort (BE). Our scheduling policy can accommodate both types of users by adapting the

weight functions W_k. In particular, the weights of the BE users are constant $W_k = 1$, while the RT users weights are functions of the DPS. As a result, the probability that a RT user is scheduled increases when its packet has a high DPS. BE users are then served when the RT users' channels are poor or the RT users' packets packet have a low DPS.

2.3.1 Comparison of Proposed Scheduler with Prior Work

To emphasize the novelty of the proposed scheduler, it is now compared with others in the literature. In [Liu et al., 2006b], the following scheduler was proposed for RT users:

$$k^* = \arg\max_k \phi_k(t), \tag{2.10}$$

where $\phi_k(t)$ is defined in terms of the instantaneous rate and the longest packet waiting time (see Section III-B in [Liu et al., 2006b] for details).

The alternative approach is based on NUM. An important example of this approach is in [Hou and Kumar, 2010], where scheduling is performed to solve the problem of maximizing the long-term rate subject to a long-term service rate constraint. In particular, the optimization problem is

$$\begin{aligned} &\text{maximize} && \sum_{k=1}^{K} U_k(q_k) \\ &\text{subject to} && \text{Network dynamics and feasibility constraints} \\ & && q_n \geq \underline{q}_k, \ \forall k, \end{aligned} \tag{2.11}$$

where q_k is the long-term service rate for user k and U_k is the long-term rate for user k.

Observe that our scheduler is most closely related to the scheduler in [Liu et al., 2006b]. The key difference is that the DPS is considered instead of the head of the line delay. This means that the proposed scheduler is simpler to implement and also that it can be characterized in terms of delay performance, BER and also packet loss due to buffer overflow, as will be shown in the remainder of the chapter.

2.4 Delay Performance

In this section, the statistics of the scheduling delay are derived for each user when there is at least one packet in each queue. In particular, the statistics of the scheduling delay of each user are given by the stationary distribution of the multidimensional Markov chain.

2.4.1 Normalized Service Rate

First, average normalized service rate is derived for the k-th user, i.e., the probability that the k-th user is scheduled. Denote $P_k(\mathbf{s})$ as the normalized service rate when the users' queue states are the elements of the state vector $\mathbf{s} = [s_1, \ldots, s_K]^T$, where each s_k, $k = 1, \ldots, K$ denotes the number of scheduling opportunities that the packet for user k has been waiting at the front of the queue. The normalized service rate for user k in state $\mathbf{s}$, $P_k(\mathbf{s})$, is then given by the following lemma.

Lemma 2.1. *The probability that user k is scheduled queue state* $\mathbf{s}$ *is given by*

$$P_k(\mathbf{s}) = \sum_{\substack{i=1 \\ i\neq k}}^{K} \frac{W_k(s_k)}{W_k(s_k) + W_i(s_i)} + \sum_{\substack{i=1 \\ i\neq k}}^{K} \sum_{m=1}^{K-2} \sum_{\substack{n_p = n_{p-1}+1 \\ p=1,\ldots,m}}^{K-2-m+p} (-1)^m \times \frac{W_k(s_k)}{W_i(s_i) + W_k(s_k) + W_i(s_i) W_k(s_k) \sum_{j=1}^{m} \tilde{W}_{n_j}^{-1}(s_{n_j})},$$

where

$$(\tilde{W}_q(s_q))_{q=1}^{K-2} = (W_p(s_p))_{p=1, p\neq i,k}^{K}. \tag{2.12}$$

Proof: See Appendix A.1. ■

Remark: In the case where each weight is equal, Lemma 1 reduces to

$$P_k = \frac{1}{K},$$

which shows that each user is equally likely to be scheduled. As a result, the normalized service rate can be interpreted as the relative priority of each user, and in special cases reduces to an extremely simple form.

2.4.2 Delay in Packet Scheduling

Next, the statistics of the DPS are derived. The probability that the current state vector is $\mathbf{s}$ is required. Denote $\mathbf{s}(n)$ as the state vector after n transmission slots. The state vectors then form a Markov chain as

$$\Pr(\mathbf{s}(n)|\mathbf{s}(1),\ldots,\mathbf{s}(n-1)) = \Pr(\mathbf{s}(n)|\mathbf{s}(n-1)). \tag{2.13}$$

Note that the transition probability from state $\mathbf{s}(n-1)$ to state $\mathbf{s}(n)$ when user k is scheduled is given by $P_k(\mathbf{s}(n-1))$. Hence, the scheduler forms a K-dimensional Markov chain with a countably infinite state space. In general, the required eigenvalue equation is intractable and it is not possible to obtain closed form expressions. As a result, the steady state characteristics are approximated by truncating the Markov chain and forming a 1-dimensional Markov chain with an augmented transition matrix. This technique for approximating the K-dimensional Markov chain is known as generating the augmented Markov chain. It has been well-studied and used in several applications such as [Liu et al., 2005]. Later via simulations it is shown that the approximation is accurate. The approximation proceeds as follows:

1. Determine the required maximum DPS for each user to achieve a given accuracy of the approximation. Denote the largest of these as d.

2. Enumerate in lexicographic order all possible state vectors with integer elements greater than or equal to one, with each element less than or equal to d.

3. Let V be the set of states that contain a single element $\mathbf{s}_k^i = 1$, where $\mathbf{s}_k^i$ is the k-th element of the i-th state vector in the lexicographic enumeration. Note that the set V can be written as

$$V = \{\mathbf{s}^i | \exists \text{ a unique } k \in \{1,2,\ldots,K\} \text{ such that } \mathbf{s}_k^i = 1\}. \tag{2.14}$$

Define S as

$$S = \{\mathbf{s}^j \in V | \mathbf{s}_m^j = \mathbf{s}_m^i + 1 \vee \mathbf{s}_m^j = \mathbf{s}_m^i = d, \forall m \neq k\},$$

where i is the index of the enumerated state for the current state vector and j is the index of the enumerated state for the future state vector. Next, the transition probability matrix $\mathbf{P}$ is constructed as

$$p_{ij} = \begin{cases} P_k(\mathbf{s}^i),\ \mathbf{s}^j \in S, \\ 0,\ \text{otherwise} \end{cases}, \tag{2.15}$$

where p_{ij} is the (i,j)-th element of $\mathbf{P}$.

4. Adjust p_{i1} such that $\sum_j p_{ij} = 1$ for all i. This is achieved by setting $p_{i1} = 1 - \sum_{j \neq 1} p_{ij}$, which ensures that $\mathbf{P}$ is a stochastic matrix. Note that for sufficiently large d, this adjustment is small.

Note that the augmented matrix $\mathbf{P}$ represents a finite 1-dimensional Markov chain, which is irreducible and ergodic. As a result, the stationary distribution π of the Markov chain exists and is unique. The steady state characteristics are obtained using the stationary distribution given by $\pi = \pi\mathbf{P}$. The stationary distribution obtained from the augmented transition matrix is an approximation for the probability of the scheduler being in any valid state $\mathbf{s}$ when the system is in steady state.

The PMF of the DPS for user k is the probability that user k is in state s_k. This is given by

$$\Pr(s_k = l) = \sum_{\substack{s_n=1 \\ n=1,\ldots,K}}^{\infty} \pi_{\mathbf{s}} \mathbf{1}_{\{s_k=l\}}, \tag{2.16}$$

where $\mathbf{1}$ is the indicator function and $\boldsymbol{ß}_{\mathbf{s}}$ is the element of the stationary distribution corresponding to state $\mathbf{s}$. In the case where the truncated Markov chain is used, the PMF of the DPS for user k is well approximated by

$$\Pr(s_k = l) \approx \sum_{\substack{s_n=1 \\ n=1,\ldots,K}}^{d} \pi_{\mathbf{s}} \mathbf{1}_{\{s_k=l\}}, \tag{2.17}$$

when d is sufficiently large. This result is crucial as it forms the basis of the SEP and PLP analysis in the sequel.

$$F_{\gamma_2|\mathbf{s}}(\gamma|\mathbf{s}) = \sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\frac{1}{W_i}\left[-e^{-\frac{\gamma}{\bar{\gamma}_2}}W_i\left(1-e^{-W_k\frac{\gamma}{\bar{\gamma}_2}/W_i}\right)+\frac{1}{\frac{1}{W_i}+\frac{1}{W_k}}\left(1-e^{-W_k\frac{\gamma}{\bar{\gamma}_2}(1/W_i+1/W_k)}\right)\right]$$
$$+\sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p}\frac{(-1)^m}{W_i}\left[-\frac{e^{-\frac{\gamma}{\bar{\gamma}_2}}}{\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}}\left(1-e^{-W_k\frac{\gamma}{\bar{\gamma}_2}\left(\frac{1}{W_i}+\sum_{j=1}^{m}\frac{1}{W_{n_j}}\right)}\right)\right.$$
$$\left.+\frac{1}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_k}+\frac{1}{W_i}}\left(1-e^{-W_k\frac{\gamma}{\bar{\gamma}_2}\left(\frac{1}{W_k}+\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\right)\right] \quad (2.18)$$

2.5 Error Performance

In this section, a new expression is derived for the SEP of the scheduled user in backlogged user queues. The expression takes the form of a single-dimensional integral that can efficiently be evaluated using numerical techniques.

2.5.1 Channel Statistics

To calculate the SEP of the scheduled user, the cumulative distribution function (CDF) of the end-to-end SNR is required. As the CDF is dependent on the state vector $\mathbf{s}$, the conditional CDF is first calculated. As detailed in Section 2.2.2, we have assumed that each channel is i.i.d. This is a reasonable assumption in heterogeneous environments (such as in cities) as each link will be affected by different factors. It is important to note that while we focus on the i.i.d scenario, our technique can also be readily extended to independent and non-identically distributed environments. The expression for the conditional CDF of the SNR of the relay-user link of the scheduled user k^* is given in Lemma 2.2. To simplify the notation, write $\gamma_2 = \gamma_{Rk^*}$ for the SNR of the scheduled user's relay-to-user link.

Lemma 2.2. *The conditional CDF of the second hop $\gamma_2 = \gamma_{Rk^*}$ is given by (2.18)*

Proof: See Appendix A.2. ■

Now, the conditional CDF of the end-to-end SNR of the source-relay-user link is derived. The conditional CDF of the end-to-end SNR, γ_{eq} in (2.5), is given in terms of the

$$F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s}) = 1 - e^{-\frac{\gamma}{\overline{\gamma}_1}} + \sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\frac{1}{W_i}\left(-W_i\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_1) + W_i\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_2) + \frac{W_iW_k}{W_i+W_k}e^{-\frac{\gamma}{\overline{\gamma}_1}}\right.$$
$$\left. - \frac{W_iW_k}{W_i+W_k}\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_3)\right) + \sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p}\frac{(-1)^m}{W_i}\left(-\frac{1}{\left(\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_4)\right.$$
$$+\frac{1}{\left(\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_5) + \frac{1}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_k}+\frac{1}{W_i}}e^{-\frac{\gamma}{\overline{\gamma}_1}}$$
$$\left.-\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}\frac{1}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_i}+\frac{1}{W_k}}G(T_6)\right), \tag{2.20}$$

source-relay and relay-user link SNRs. Using a result in [Yang et al., 2011], the conditional CDF of the end-to-end SNR may be written as

$$F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s}) = 1 - \int_0^\infty \left(1 - F_{\gamma_2|\mathbf{s}}\left[\gamma + \frac{\gamma^2+c\gamma}{\omega}|\mathbf{s}\right]\right)$$
$$\times f_{\gamma_1|\mathbf{s}}(\omega+\gamma|\mathbf{s})d\omega, \tag{2.19}$$

where $\gamma_1 = \gamma_{SR}$.

Using (2.19), the conditional CDF of the end-to-end SNR conditioned on the state vector $\mathbf{s}$ is given in Lemma 2.3.

Lemma 2.3. *The conditional CDF of the end-to-end SNR conditioned on* $\mathbf{s}$ *is given by (2.20) where*

$$G(T_i) = e^{-\gamma T_i}2\sqrt{\overline{\gamma}_1 T_i(\gamma^2+c\gamma)}K_1\left(2\sqrt{\frac{T_i(\gamma^2+c\gamma)}{\overline{\gamma}_1}}\right),$$

with $T_1 = T_4 = \frac{1}{\overline{\gamma}_2}$, $T_2 = \frac{1}{\overline{\gamma}_2}\left(1+\frac{W_k}{W_i}\right)$, $T_3 = \frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i}+\frac{1}{W_k}\right)$, $T_5 = \left(\frac{1}{\overline{\gamma}_2}+\frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i}+\sum_{j=1}^{m}\frac{1}{\tilde{W}_{n_j}}\right)\right)$, $T_6 = \frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_k}+\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)$, *and* K_1 *is the first order modified Bessel function of the second kind.*

Proof: See Appendix A.3. ■

The average CDF of the end-to-end SNR is presented in Theorem 2.1, which is ob-

tained by summing over the result in Lemma 2.3.

Theorem 2.1. *The average CDF of the end-to-end SNR is given by*

$$F_{\gamma_{eq}}(\gamma) = \sum_{s_1,\ldots,s_K} F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s})\pi_{\mathbf{s}},$$

where $F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s})$ *is given in (2.20) and* $\pi_{\mathbf{s}}$ *is the probability that the DPS state vector is* $\mathbf{s}$.

2.5.2 Symbol Error Probability

Following [Yang et al., 2011], the SEP of the scheduled user for different modulation formats can be evaluated according to

$$P_S = \frac{a}{2}\sqrt{\frac{b}{\pi}}\int_0^\infty \gamma^{-\frac{1}{2}} e^{-b\gamma} F_{\gamma_{eq}}(\gamma)\, d\gamma. \tag{2.21}$$

The constants a and b are modulation-specific with $a = 1,\ b = 1$ for BPSK, $a = 1,\ b = 0.5$ for QPSK, and $a = 2\left(M-1\right)/M,\ b = 3/\left(M^2-1\right)$ for M-PAM.

Note that (2.21) is absolutely convergent. As such, the sum in Theorem 2.1 and the integral in (2.21) can be swapped by applying the dominated convergence theorem. This ensures that the infinite sum converges. The integral can then be evaluated efficiently using numerical integration, leading to reduced evaluation time compared with Monte Carlo simulation. The analytical approach is shown to accurately approximate the Monte Carlo simulation.

2.6 Packet Loss Performance

In this section, the PLP of each queue is analyzed using the proposed scheduling policy. This is achieved by constructing a new Markov chain for the buffer states for each queue with transition probabilities dependent on the scheduling policy, arrival rate, and transmission time.

2.6.1 Buffer State

The PMF of the buffer state that gives the probability that the buffer has l, $0 \leq l \leq B$ packets is now obtaineds. Note that the buffer state is measured at the beginning of a scheduling slot, after a packet is scheduled in the current slot, and before new arrivals. This is important as the time when the buffer state is measured affects the PMF of the buffer state and subsequently the PLP. Also note that the buffer state is independent of the DPS corresponding to user k. To calculate the PMF of the buffer state, the average probability that user k is scheduled is required, which is given by

$$P_k = \sum_{\substack{s_n=1 \\ n=1,\ldots,K}}^{\infty} P_k(\mathbf{s})\pi_{\mathbf{s}} \tag{2.22}$$

where $P_k(\mathbf{s})$ is the probability user k is scheduled in DPS state $\mathbf{s}$ and $\pi_{\mathbf{s}}$ is the probability that the DPS state vector is $\mathbf{s}$. An accurate approximation of (2.22) can be obtained using the stationary distribution arising from the truncated multidimensional Markov chain given in (2.17).

To obtain the PMF of the buffer state, the stationary distribution of the associated Markov chain is required. Since the buffer state of each user is only dependent on the individual user's statistics, the Markov chain is one-dimensional and the stationary distribution can be obtained by explicitly constructing the transition matrix. In particular, the construction of the transition matrix for user k, $\mathbf{T}_k$, is shown in Table 1. In the table, t_{ij} is the (i,j)-th element of $\mathbf{T}_k$ where $1 \leq i,j \leq B+1$. The (i,j)-th element represents the transition from the buffer storing $i-1$ packets to the buffer storing $j-1$ packets. Note that since the buffer is finite, the corresponding irreducible and ergodic Markov chain is also finite. As a result, it has a unique stationary distribution.

In order to analyze the PMF of the buffer state, assume that it is possible for a user to be scheduled without a packet. This is necessary as the weight in the proposed scheduling policy does not account for the buffer state. Simulations show that the approximation is accurate in spite of this assumption. The PMF of the buffer state approximation for user k is obtained by solving the eigenvalue equation $\nu_{B,k} = \nu_{B,k}\mathbf{T}_k$, which is solved by finding the left eigenvector of $\mathbf{T}_k$ corresponding to an eigenvalue of one. Note that if the schedul-

Table 2.1: Construction of transition matrix $\mathbf{T}_k$

Case 1: $j = i$ i) if $j = 1$, then $t_{ij} = \Pr(N_k(T) = 0) + \Pr(N_k(T) = 1)P_k$ ii) if $j < B$, then $t_{ij} = \Pr(N_k(T) = 0)(1 - P_k)\Pr(N_k(T) = 1)P_k$ iii) if $j = B$, then $t_{ij} = (1 - P_k)\Pr(N_k(T) = 0) + (1 - \Pr(N_k(T) = 0))P_k$
Case 2: $j = i - 1$ i) if $i < B + 1$, then $t_{ij} = \Pr(N_k(T) = 0)P_k$ ii) if $i = B + 1$, then $t_{ij} = P_k$
Case 3: $j < i - 1$ i) $t_{ij} = 0$
Case 4: $j > i$ i) if $j < B$, then $t_{ij} = \Pr(N_k(T) = j - i)(1 - P_k) + \Pr(N_k(T) = j - i + 1)P_k$ ii) if $j = B$, then $t_{ij} = \Pr(N_k(T) = j - i)(1 - P_k) + (1 - \Pr(N_k(T) \leq B - (i - 1) - 1))P_k$
Case 5: $j = B + 1$ i) if $i < B + 1$, then $t_{ij} = (1 - \Pr(N_k(T) \leq B - (i - 1) - 1)(1 - P_k)$ ii) if $i = B + 1$, then $t_{ij} = 1 - P_k$

ing weights are constants independent of the DPS, then the stationary distribution $\nu_{B,k}$ provides the exact PMF of the buffer state. This is due to the scheduler not requiring the current DPS of each packet.

2.6.2 Packet Loss Probability

The PLP is the probability that a packet is lost due to buffer overflow. Before evaluating the PLP for a given packet, the following lemma [Daley and Vere-Jones, 2003, pg. 22] is required.

Lemma 2.4. *Let the packet arrivals follow a homogeneous Poisson process. Then, the unordered packet arrival times are i.i.d uniformly on $(0, T)$ when conditioned on the number of arrivals in the interval.*

The PLP for each user can now be obtained for a given buffer size by considering the probability that the buffer is full at time $0 < t < T$ after a scheduling opportunity. Here, t is the the time of the new packet arrival. Theorem 2.2 gives an approximation of the PLP. The approximation arises due to dependence on the stationary distribution $\nu_{B,k}$ and is exact when the scheduling policy weights are fixed constants.

Theorem 2.2. *The PLP for user k is given by*

$$P_{L,k} \approx \sum_{l=0}^{B} \nu_l \left(1 - \sum_{m=0}^{B-l-1} \frac{\lambda_k^m}{Tm!} \times \left[\frac{m!}{\lambda_k^{m+1}} - e^{-\lambda_k T} \sum_{n=0}^{m} \frac{m!}{n!} \frac{T^n}{\lambda_k^{m-n+1}} \right] \right) \tag{2.23}$$

where ν_l is the element of the stationary distribution corresponding to a buffer state l for user k. Moreover, $P_{L,k}$ is exact when the weights $W_k(s_k),\ k = 1, \ldots, K$ are constants.

Proof: See Appendix A.4. ■

The PLP approximation shows the clear dependence on the the transmission time and arrival rates for the user under consideration. Intuitively, if the arrival rate is high or the transmission time long, the PLP due to buffer overflow is large. A consequence of

this is that additional redundancy through channel coding does not always improve the throughput.

Note that the expected total packet delay can be obtained via Little's law [Zalesky, 2009] from the buffer state distribution and the PLP. In particular,

$$E[W_k] = L_k / \lambda_{e,k}, \tag{2.24}$$

where $E[W_k]$ is the expected total packet delay for user k, L_k is the expected number of packets in the queue for user k, which can be obtained using (20), and $\lambda_{e,k}$ is the effective arrival rate for user k given by

$$\lambda_{e,k} = \lambda_k(1 - P_{L,k}), \tag{2.25}$$

where λ_k is the actual packet arrival rate and $P_{L,k}$ is the PLP of the user given by (21). Surprisingly, this means that the expected total packet delay can be quantified, despite only exploiting the DPS in the proposed scheme.

In the next section, numerical and simulation results are used to evaluate the effect of scheduling policy and examine the accuracy of the approximation when the weights are not fixed.

2.7 Simulation and Numerical Results

2.7.1 DPS and SEP

The analytical results for the PMF of the DPS and the SEP of the scheduled user are now confirmed using Monte Carlo simulations. Simulations are performed with distances $d_S = d_R = 1$ and transmit powers $E_S = E_R = 1$. The WMRN consists of a single RT user and $K - 1$ BE users. To illustrate the behavior of the proposed scheduler, consider the following exponential-based scheduling policies for the weights of the RT user:

$$W_1 = e^{0.07(j_I - 1)s_1} \tag{2.26}$$

for $K = 3$, and

$$W_1 = e^{0.06(j_I - 1)s_1} \tag{2.27}$$

for $K = 4$, where $j_I \in \{1, 2, \ldots\}$ is the scheduling policy index that is a reference number for each policy. The weights are exponential functions of s_1, which is the number of scheduling opportunities that the packet for the RT user has been waiting at the front of the queue. The weights of the BE users are set to $W_i = 1,\ i = 2, \ldots, K$ that means there are no delay constraints.

In Fig. 2.2, the probability that the DPS target of $\Pr(s_1 \leq 3)$ is satisfied for the RT user is plotted against the scheduling policy index. Observe the accuracy of the approximation in (2.17) by noting that the simulations curves are consistent with the analytical curves. The figure shows that the probability of satisfying the DPS target is greater than 90% for $j_I \geq 5$ when $K = 3$ and for $j_I \geq 8$ when $K = 4$. When $K = 3$, see that the adaptive weight policy of $j_I = 6$ provides a 30% improvement in satisfying the DPS target relative to the equal weight policy of $j_I = 1$, which was considered in [Yang et al., 2011]. This improvement is indicated in the figure by the arrow. Also observe that the probability of satisfying the DPS target increases as the scheduling policy index increases. The effect of this on the SEP of the scheduled user is further examined in Figs. 2.3 and 3.

In Fig. 2.3, the SEP of the scheduled user (i.e., k^* in (2.8)) is plotted against the average SNR $\overline{\gamma}_1$. The SEP of the scheduled user for $K = 3$ is plotted with the adaptive weight policy of $j_I = 6$ in (2.26) and the equal weight policy in of $j_I = 1$ in (2.26). For comparison purposes, the SEP of the scheduled user is also plotted for a short delay policy where $W_1(s_1) = 100$. The short delay policy closely approximates a single user relay network where the probability of satisfying the DPS target is greater than 99% (evaluated using (2.17)). The figure shows that the SEP of the adaptive weight policy and the equal weight policy outperform the short delay policy by 2 dB at the SEP of 10^{-2}. Also observe that the adaptive weight policy provides a comparable match to the equal weight policy. This means that the 30% improvement in the probability of satisfying the DPS target of the RT user in Fig. 2.2 is attained with negligible degradation in the SEP of the scheduled user in Fig. 2.3.

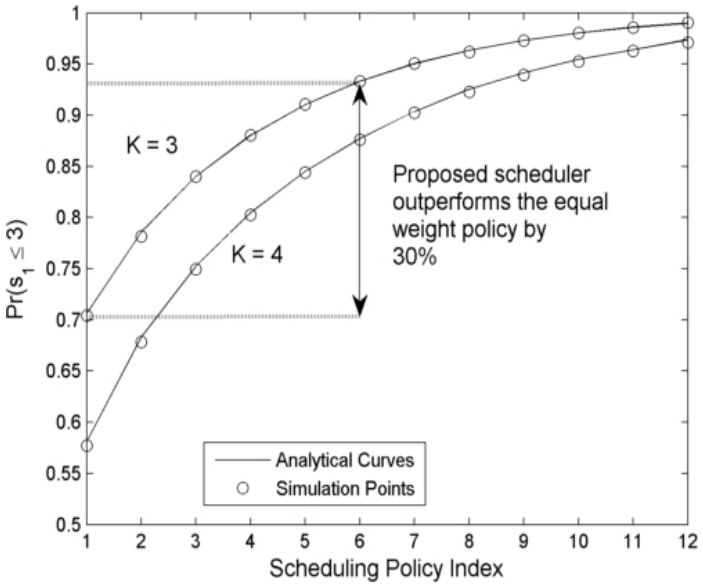


Figure 2.2: The probability the DPS target of $s_1 \leq 3$ is satisfied for the RT user versus the scheduling policy index of several policies.

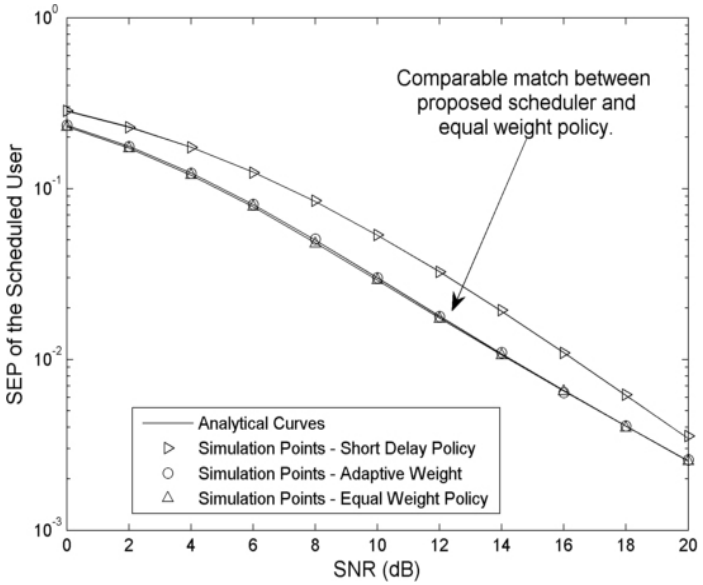


Figure 2.3: The SEP of the scheduled user (k^* in (2.8)) versus the SNR using BPSK with $K = 3$. Three scheduling policies are considered: short delay, adaptive weight and equal weight.

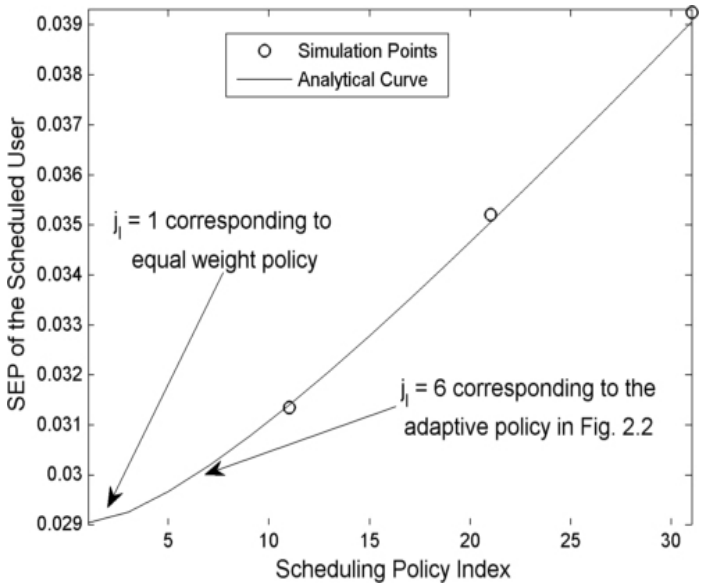


Figure 2.4: The SEP of the scheduled user at 10 dB versus the scheduling policy index with $K = 3$ using BPSK.

In Fig. 3, the SEP of the scheduled user at $\overline{\gamma}_1 = 10$ dB is compared with the scheduling policy index in (2.26) for $K = 3$. Observe that an increase in the scheduling policy index causes an increase in the SEP of the scheduled user. Moreover, observe that the SEP for the adaptive policies of $j_I \leq 6$ is less than 4% greater than the SEP for the equal weight policy of $j_I = 1$. This is consistent with the result in Fig. 2 that showed a comparable match between the SEP of the scheduled user using policies $j_I = 6$ and $j_I = 1$.

2.7.2 PLP

In Fig. 4, the PLP with $K = 3$ is plotted for varying buffer sizes. The arrival rates are set as $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$. Two scheduling policies are plotted: an adaptive policy where $W_1(s_1) = e^{0.2s_1}$, $W_2(s_2) = e^{0.1s_2}$ and $W_3 = 1$, and the equal weight policy [Yang et al., 2011]. Observe that analytical and simulation results are in good agreement for all users and scheduling policies. As expected, the users with higher priorities (i.e., users 1 and 2 of the adaptive policy) have a lower PLP. For $B > 7$, observe that the PLP is independent of the buffer size for all users with $T = 5$ ms. This is due to the fact that when $T = 5$ ms, the average scheduling rate P_k/T for each user is less than the average

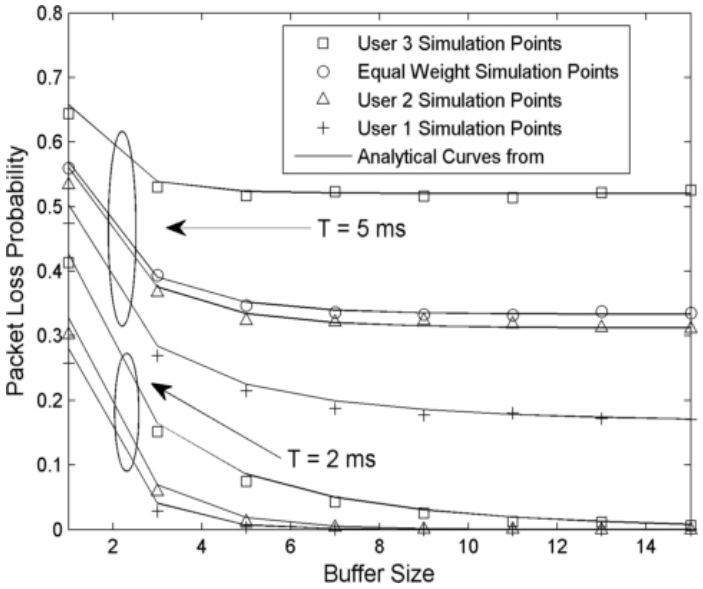


Figure 2.5: The PLP for each user versus the buffer size with $K = 3$, transmission times $T = 5$ ms and $T = 2$ ms and arrival rate $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$. Two scheduling policies are considered: equal weight ($W_i = 1$, $i = 1, 2, 3$) and adaptive ($W_1 = e^{0.2s_1}$, $W_2 = e^{0.1s_2}$, $W_3 = 1$).

arrival rate $\lambda_k = 0.1$ ms^{-1}, $\forall k$. As a result, the queue experiences a positive drift and the buffer overflows regardless of the length. In contrast, the PLP of each user approaches zero with increasing buffer size for $T = 2$ ms. This is consistent with the results for queue stability in the infinite buffer scenario (see for example, Lemma 1 in [Neely et al., 2002, Yeh and Cohen, 2003]). This means that the transmission time should be minimized to improve the service rate and allow large buffer sizes to be exploited.

To obtain further insights, the PLP of the RT user with the adaptive weight policy considered in Fig. 2.3 is compared to the equal weight policy [Yang et al., 2011]. The PLP of the RT user with the adaptive weight policy of $j_I = 6$ in (22) is evaluated using (21) as $P_L = 0.04$, for a buffer size $B = 15$, transmission time $T = 5$ ms, and arrival rate $\lambda_k = 0.1$ ms $k = 1, 2, 3$. Observe that this results in a greater than a 85% improvement over the PLP of the equal weight policy plotted in Fig. 4. This shows that the proposed scheduler can achieve a significant improvement in the PLP compared with the equal weight scheduler, with negligible degradation in the SEP of the scheduled user.

2.7.3 Transmission Time

In Fig. 2.5 it was observed that increasing the transmission time impacts on the PLP. To determine the optimal transmission time, the effect of channel coding must be accounted for. Of course, when employing coding, a longer transmission time is required to account for the redundancy in the signal. To examine the tradeoff between the coded SEP of the scheduled user and the PLP for each queue, the throughput is considered

$$\text{Throughput} = (1 - P_{E,coded})(1 - P_{L,ave}), \tag{2.28}$$

where $P_{E,coded}$ is the coded SEP of the scheduled user and $P_{L,ave}$ is the average PLP over all queues. The average PLP over all the queues is given by

$$P_{L,ave} = \frac{1}{K} \sum_{i=1}^{K} P_{L,i}. \tag{2.29}$$

The throughput expression in (2.28) approximates the WMRN as a single point-to-point link using a single queue with a PLP given by the average over all queues. As a result, (2.28) gives a simple characterization of a WMRN as transmission times are varied.

In Fig. 2.6, the throughput is compared to the inverse of the code rate R^{-1}, for varying uncoded transmission time T_{unc}. Here $T = T_{unc}R^{-1}$, where R is the normalized rate of the coded signal and the rate of the uncoded signal is $R = 1$. The data is coded using punctured convolutional codes. The distances and transmit powers are $d_S = d_R = 1$ and $E_S = E_R = 1$. In the simulation, the buffer size is $B = 20$, the arrival rate is $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$, the number of transmitted QPSK data symbols is 100, and the scheduling policy is $W_1(s_1) = e^{0.2s_1}$, $W_2(s_2) = e^{0.1s_2}$, $W_3 = 1$.

In the figure observe that the throughput increases with increasing transmission time, T, when $T_{unc} = 0.5$ ms. In contrast, the throughput decreases with increasing T when $T_{unc} = 2$ ms. Of particular interest is the scenario where T_{unc} is between 0.5 ms and 2 ms. From the figure observe that when $T_{unc} = 1$ ms, the throughput does not vary monotonically with T. This suggests that an efficient tradeoff between the PLP and the coded SEP of the scheduled user for WMRNs with finite buffers is critically dependent

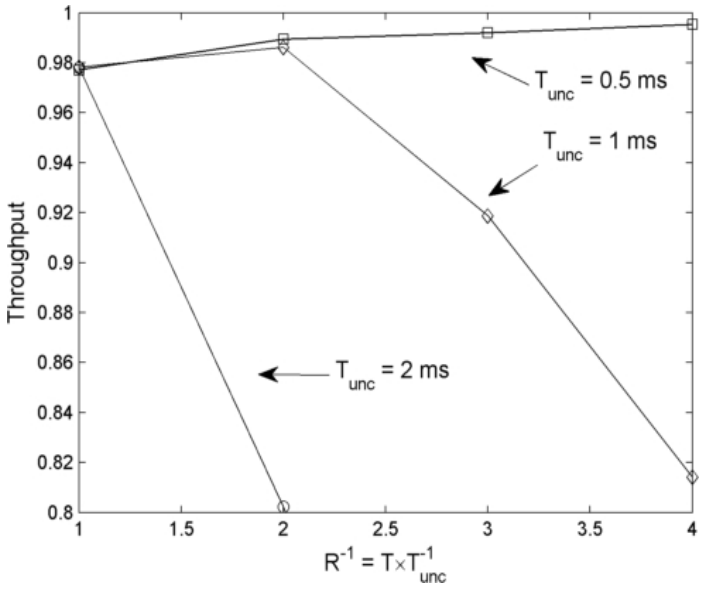


Figure 2.6: The throughput of an equivalent single user network versus the inverse of the code rate R^{-1} for varying uncoded transmission times T_{unc}, arrival rate $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$, and scheduling policy ($W_1 = e^{0.2s_1}$, $W_2 = e^{0.1s_2}$, $W_3 = 1$)

on T_{unc}. This leads to the following key design rules: 1) in the small T_{unc} regime, the code rate should be minimized, 2) in the large T_{unc} regime, the code rate should be maximized and 3) in the intermediate T_{unc} regime, the code rate should lie between $0 < R < 1$.

2.8 Conclusion

A new user scheduler for WMRNs was proposed. A cross-layer scheduling approach was considered that accounted for delay in packet scheduling (DPS), symbol error probability (SEP), and packet loss probability (PLP) due to buffer overflow. The user with the largest weighted SNR was scheduled, where the weight was a function of the DPS. A new analytical expressions was derived for the DPS and the SEP of the scheduled user. Analysis and simulation was provided to confirm the performance of the scheduler. The probability that a target DPS is met was shown to be 30% higher for the proposed new scheme compared to the standard opportunistic equal weight scheduler, with negligible degradation in the SEP of the scheduled user. Finally, channel coding was exploited and the transmission time to optimize the tradeoff between the PLP, and the SEP of the

scheduled user. This revealed new design rules for the transmission time when coding is used.

Chapter 3

Cross-layer Scheduling in Coordinated Multipoint

In this chapter, a CoMP downlink scheduler is proposed to reduce the packet loss probability (PLP) due to buffer overflow in BSs with finite queues. The proposed CoMP scheduler selects a single BS to serve the associated cell-edge user with the largest weighted SNR. To meet PLP targets, a simple strategy is developed to design the packet transmission time and the scheduling weights of each BS. The network design capitalizes on a new closed-form expression for the PLP that relates three key network parameters: packet arrival rate, packet transmission time, and probability that each BS is scheduled.

3.1 Introduction

COMP is a key technology for interference mitigation in LTE Release 10 [Irmer et al., 2011]. The core concept is to allow multiple cooperating BSs to share information via backhaul links to boost the signal-to-noise ratio (SNR) of cell-edge users. Such information may include user data, channel state information, queue length, and packet delay. One form of CoMP in the downlink is coordinated scheduling, where each user's data is only available at a single serving BS, and all the BSs cooperatively schedule transmissions [Irmer et al., 2011]. While CoMP with coordinated scheduling is effective in reducing inter-cell interference, it increases packet delay as each BS must wait its turn to transmit. As a consequence, the packet loss probability (PLP) due to buffer overflow at each BS must be accounted for in the CoMP scheduler design.

The CoMP scheduler design can be viewed as a generalization of the classical downlink scheduler design. The main difference is that the queue and channel information

used for scheduling in CoMP is limited by the finite bandwidth of the backhaul links. In the classical downlink scheduler design, the two main approaches are network utilization maximization (NUM) and heuristic scheduling. In NUM, an optimization problem is solved with a minimum service rate constraint [Hou and Kumar, 2010, O'Neill et al., 2008]. Unfortunately, the high computation complexity of NUM scheduling, which typically requires iterative algorithms, may not be practical in CoMP networks. Alternatively, heuristic scheduling uses low complexity policies such as the weighted SNR (or sum-rate) policy [Kobayashi and Caire, 2006, Egan et al., 2013b, Liu et al., 2006b], where the user with the largest weighted SNR is scheduled. The weight is typically chosen to be a function of the packet delay. While this approach has a low computational complexity, the high feedback overhead required to constantly update the scheduler weights is prohibitive in CoMP networks with limited backhaul bandwidth. Heuristic scheduling in CoMP networks that consider the user throughput have been proposed in [Zhang and Andrews, 2010, Zhang et al., 2011, Seifi et al., 2011].

In this chapter, a cross-layer approach is proposed for a heuristic CoMP scheduler that addresses the user throughput as well as the PLP in practical networks with finite queue lengths and limited backhaul bandwidth between the BSs. In the proposed scheduler, the BS serving the cell-edge user with the maximum weighted SNR is selected to transmit. It is important to note that the cell-edge users are served on different sub-carriers from the inner cell users, which means that the cell-edge user scheduling does not affect the inner cell users. The weights are designed to be fixed, which means that the only information that needs to be exchanged between the BSs is each user's SNR. As such, the backhaul requirements are reduced compared with adaptive weight schedulers that utilize instantaneous packet delay information.

The proposed CoMP scheduler design consists of two key components: scheduling policy weights and packet transmission time. First, a new upper bound is derived for the PLP due to buffer overflow. The analytical expression for the PLP bound is in closed-form, in contrast with previous approaches that are semi-analytical and require the stationary distribution of large Markov chains [Egan et al., 2013b]. The PLP expression relates three key network parameters: packet arrival rate, packet transmission time, and

probability that each BS is scheduled. Based on this a new strategy is proposed to design the packet transmission time and the scheduling weights of each BS.

To compare the PLP under different heuristic scheduling policies, the τ-achievable region is introduced, which is the region of achievable PLPs for transmission time $T \geq \tau$. The τ-achievable region explicitly shows whether a given scheduling policy can achieve the target PLP for each BS at a given transmission time T. A sufficient condition is also obtained for queue stability in the case of infinite buffers. Using this condition, the maximum transmission time is compared for queues with buffer size $B = 1$ and queues with infinite buffer capacity, which results in lower and upper bounds on the maximum transmission time, respectively. The proposed fixed weight scheduler is compared with the adaptive weight scheduler that requires instantaneous packet delay information [Egan et al., 2013b, Liu et al., 2006b, Liu et al., 2006a]. Most importantly, analysis and simulation show that the proposed fixed weight scheduler can achieve a comparable PLP to the adaptive weight scheduler, while reducing communication overheads between the BSs.

3.2 System Model

Consider a coordinated single-cell transmission network where K BSs service K cell edge users, each user corresponding to a different BS; illustrated in Fig. 3.1. This scenario can occur when cell-edge users share their own OFDMA subcarriers to improve throughput. Each BS is equipped with a first-in first-out (FIFO) data queue, which stores data corresponding to a given cell edge user. To balance the PLP and throughput requirements, a cross-layer scheduling policy is proposed, as outlined in Section 3.3.1. The proposed policy jointly accounts for throughput via the SNR and user priority via the scheduling weights.

It is important to note that the design choice that cell-edge users share their own OFDMA subcarriers is intimately linked to the architecture of the CoMP network. CoMP networks are the first to allow cooperation in physical layer transmission between cells. When joint transmission is allowed between all users in each cell, it also means that the MAC layer design must also jointly consider all users in multiple cells. This significantly

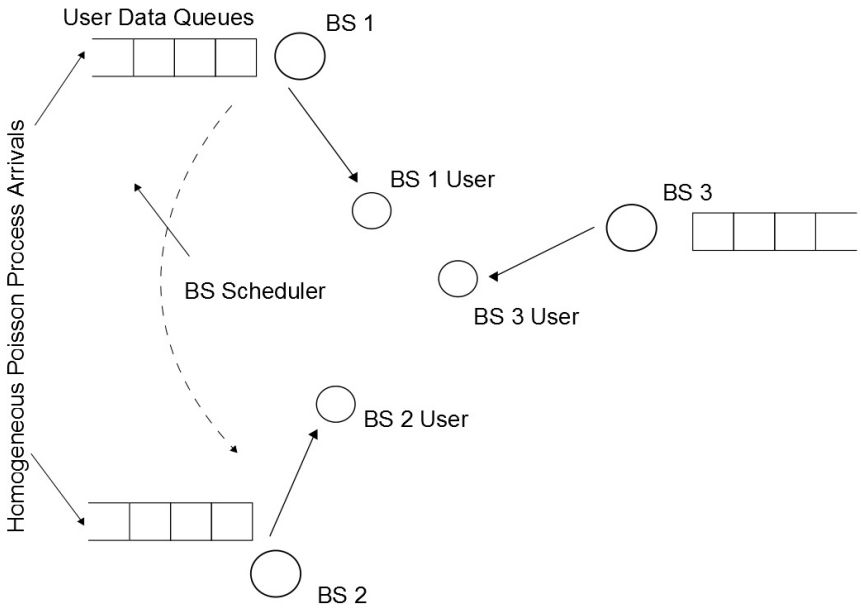


Figure 3.1: Illustration of the CoMP network.

complicates the design and makes it more difficult to ensure MAC layer metrics such as queue stability and packet delay satisfy the desired constraints. As such, it is highly desirable to only allow joint transmission for cell-edge users as this means that service to these users can be improved, while ensuring that the network architecture is not significantly modified.

3.2.1 MAC Layer Architecture

Each BS has a finite queue with buffer size B, where each BS corresponds to a distinct user. A user's packet is lost if the buffer for the queue at its BS is full and a new packet arrives. The arrival process of the packets for each queue is assumed to be a homogeneous Poisson process with rate λ_k, $k = 1, \ldots, K$, where each k corresponds to a different BS. Assume that each packet has a fixed size. As such, the probability that n packets arrive in an interval of time T for the k-th user is then given by

$$\Pr(N_k(T,0) = n) = \frac{e^{-\lambda_k T}(\lambda_k T)^n}{n!}. \tag{3.1}$$

Packets can be re-requested with the caveat that the arrival of the re-requested packet is consistent with the Poisson arrival process. The transmission time T is the same for all BSs.

3.2.2 PHY Layer Architecture

Consider a block fading channel for each user, with a coherence time of T_c. Each BS transmits over a time slot of T seconds, where $T \leq T_c$. At user k, the received signal is given by

$$y_k = \sqrt{E_S d^{-\eta}} h_k x + z_k, \tag{3.2}$$

where E_S is the transmit power at the source, h_k is the Rayleigh fading channel coefficient between the BS and user k, x is the transmitted symbol, and z_k is the additive white Gaussian noise (AWGN) with one-sided power spectral density N_0. The instantaneous signal-to-noise ratio at user k is given by

$$\gamma_k = \frac{d^{-\eta} E_S |h_k|^2}{N_0}, \tag{3.3}$$

where the effect of path loss is reflected through the distance d and the path loss exponent η. Assume that the distance d is the same at each user. As each user experiences i.i.d Rayleigh fading, the probability density functions (PDFs) of the instantaneous SNRs are

$$f_{\gamma_k}(\gamma) = \frac{1}{\overline{\gamma}} e^{-\frac{\gamma}{\overline{\gamma}}},\ k = 1, 2, \ldots, K. \tag{3.4}$$

where $\overline{\gamma} = E[\gamma_k],\ \forall k \in \{1, \ldots, K\}$, is the average SNR in each user's link, with $E[\cdot]$ denoting the expectation.

3.3 Fixed Weight CoMP Scheduling Policies

In this section, the proposed fixed weight scheduling policy is detailed. The aim is to satisfy the PLP of the K cooperating BSs, which means that the weights are functions of

the packet arrival rates and the transmission time for each user. Fixed weight policies are desirable as instantaneous queue information is not required at any BS, which in turn reduces the required backhaul. A closed form expression is derived for the PLP due to buffer overflow for queues with a buffer size $B = 1$. This provides a useful upper bound on the PLP for arbitrary buffer sizes. The τ-achievable region is then introduced, which represents the achievable PLPs for each user with transmission time $T \leq \tau$. The τ-achievable region is treated in detailed for $K = 2$ BSs in detail.

3.3.1 CoMP BS Scheduler

The proposed CoMP BS scheduler selects the user with the largest weighted SNR with *fixed*, and potentially different, weights allocated to each BS. The weights are restricted such that each weight is fixed is an important design choice to reduce the required backhaul overhead of the links between the cooperating BSs.

The proposed CoMP scheduling policy is given by

$$k^* = \arg \max_{k=1,\ldots,K} \gamma_k W_k, \tag{3.5}$$

where

(i) k^* is the scheduled user;

(ii) γ_k is the instantaneous SNR of the link between the BS and the user, where $k = 1, \ldots, K$;

(iii) W_k is the weight that satisfies

$$W_k \geq 0,\ k = 1, \ldots, K \tag{3.6}$$

where W_k is a constant weight for the k-th BS.

It is important to note the similarity with the proposed scheduler in Chapter 2. The key difference between the scheduler in Chapter 2 and the scheduler in (3.5) is that the former deals with scheduling users in a single BS network, while the later deals with the

CoMP network with multiple BSs. As such, the related work detailed in Section 2.3.2 can also be compared with the proposed scheduler in (3.5).

3.3.2 Packet Loss Probability Analysis

To calculate the PLP bound, the probability that user k is scheduled in a multiple BS network with Rayleigh fading channels is first obtained. This corresponds to the normalized service rate of user k. With the help of [Egan et al., 2013b], this is presented in the following lemma.

Lemma 3.1. *The probability that user k is scheduled is given by*

$$P_k = \sum_{\substack{i=1 \\ i\neq k}}^{K} \frac{W_k}{W_k + W_i} + \sum_{\substack{i=1 \\ i\neq k}}^{K} \sum_{m=1}^{K-2} \sum_{\substack{n_p = n_{p-1}+1 \\ p=1,\ldots,m}}^{K-2-m+p} (-1)^m \times \frac{W_k}{W_i + W_k + W_i W_k \sum_{j=1}^{m} \tilde{W}_{n_j}^{-1}},$$

where

$$(\tilde{W}_q(s_q))_{q=1}^{K-2} = (W_p(s_p))_{p=1,p\neq i,k}^{K}. \tag{3.7}$$

The following intuitive corollary is obtained directly from Lemma 3.1 for $K = 2$.

Corollary 3.1. *If $K = 2$, then P_k in Lemma 3.1 simplifies to*

$$P_k = \frac{W_k}{W_k + W_i}, \tag{3.8}$$

where $i, k = 1, 2,\ i \neq k$.

Next, the distribution of the time between the last scheduling opportunity and the current packet arrival time is determined. With the help of eqn 2.1.7 in [Daley and Vere-Jones, 2003], this is presented in the following lemma.

Lemma 3.2. *Let t be the packet arrival time. Then, the time between the last scheduling opportunity and t is uniformly distributed on $(0, T)$.*

To derive the new bound for the PLP due to buffer overflow, consider the scenario that a given packet arrives at time t after a scheduling opportunity. There are two cases where a packet will be dropped:

1. The buffer is full at the end of the previous time slot;
2. The buffer is not full at the end of the previous time, but at least one packet has arrived before time t.

Note that the buffer state evolves according to a two-state Markov chain. This is due to the fact that the state of the buffer at the beginning of a given time slot only depends on the buffer state at the beginning of the previous time slot. The transition probability matrix is given by

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \tag{3.9}$$

where

$$\alpha = 1 - e^{-\lambda_k T} - (1 - e^{-\lambda_k T}) P_k, \tag{3.10}$$

$$\beta = P_k, \tag{3.11}$$

and P_k is obtained from Lemma 3.1. The stationary distribution $^\circ = [\nu_0, \nu_1]$ is found using the standard result (Example 1.1.4 [Norris, 1997])

$$\nu_0 = \frac{\beta}{\alpha + \beta}, \quad \nu_1 = \frac{\alpha}{\alpha + \beta} \tag{3.12}$$

Let t be the arrival time of a given packet after the last scheduling opportunity, B_t be the event that at least one packet arrives between the last scheduling opportunity and time t, and B_l, $l = 0, 1$ be the event that the buffer of size $B = 1$ is in state l at the beginning of the time slot. The bound for the PLP for user k is then given by

$$P_{L,k} \leq \Pr(B_t)$$

$$= \Pr(B_t|B_0)\nu_0 + \nu_1. \tag{3.13}$$

Observe that $\Pr(B_t|B_0)$ is the probability that at least one packet arrives after the last scheduling opportunity and before time t. As such,

$$\begin{aligned} P_{L,k} &\leq \frac{\beta}{\alpha+\beta}\int_0^T \frac{1}{T}\Pr(N_k(t,0) > 0|t)dt + \frac{\alpha}{\alpha+\beta} \\ &= \frac{\beta}{\alpha+\beta}\int_0^T \frac{1}{T}\left(1 - e^{-\lambda_k t}\right)dt + \frac{\alpha}{\alpha+\beta}. \end{aligned} \tag{3.14}$$

Note that the $B = 1$ case provides an upper bound on the PLP as increasing the buffer size only reduces the PLP. The previous discussion is summarized in the following proposition, which provides a closed-form bound for the PLP. In particular, (3.15) is obtained by solving the integral in (3.14).

Proposition 3.1. *For buffer size $B \geq 1$, the PLP for user k satisfies*

$$\begin{aligned} P_{L,k} \leq{}& \frac{P_k}{P_k + (1-P_k)(1-e^{-\lambda T})}\frac{1}{T}\left(T + \frac{1}{\lambda_k}e^{-\lambda_k T} - \frac{1}{\lambda_k}\right) \\ &+ \frac{(1-P_k)(1-e^{-\lambda_k T})}{P_k + (1-P_k)(1-e^{-\lambda_k T})} \end{aligned} \tag{3.15}$$

The result in Proposition 3.1 is verified using Monte Carlo simulations in Section 3.5.

3.3.3 Achievable Packet Loss Probabilities

Next, the achievable packet loss probability region is defined for the network. This is used to compare the PLPs of the multiple BSs as the scheduling weights are varied. Formally, the achievable PLP region is defined as follows.

Definition 3.1. *A K-tuple of PLPs $(P_{L,1}, \ldots, P_{L,k})$ is said to be τ-achievable if there exists a scheduling policy $\mathcal{P}$ and transmission time $T \geq \tau$ pair $(\mathcal{P}, T)$ that provides the PLP K-tuple.*

The τ-achieveable PLP region offers new interesting insights when the scheduling policies adopt fixed weights. The τ-achievable region in this case is denoted as the τ_{FW}-achievable region. Note that the PLP monotonically increases with the transmission time.

As such, the boundary of the τ_{FW}-achievable region occurs when $T = \tau$. A figure providing an example of the τ-achievable region for a different scheduling policies is provided in Fig. 3.4.

The τ_{FW}-achievable region is now described in Theorem 3.1, which follows from the analysis in Section 3.3.2.

Theorem 3.1. *If each element in the PLP K-tuple $(P_{L,1}, \ldots, P_{L,K})$ satisfies the bound in (3.14) for $T = \tau_{FW}$, then the PLP K-tuple is τ_{FW}-achievable irrespective of the buffer size.*

3.4 Network Design

In this section, a simple design strategy is proposed to ensure that throughput and PLP targets are met. The proposed strategy obtains the scheduler weights and the packet transmission time, based on the $B = 1$ PLP bound in Proposition 3.1. Unfortunately, the exact PLP for $B > 1$ is obtained using semi-numerical techniques [Egan et al., 2013b]. As such, queue stability is also considered, which is analogous to the PLP for infinite buffer queues. The transmission time is compared when based on the PLP bound and queue stability to determine the difference between the small buffer regime and the large buffer regime.

3.4.1 Design Strategy

The proposed design strategy is based on the observation that the PLP of a BS is determined by the probability of scheduling, packet arrival rate, and the transmission time. As such, the transmission time can be obtained when the probability of scheduling and the packet arrival rate are known. Note that the probability that a BS is scheduled determines its user's throughput. This is due to the fact that the proportion of time slots a BS is allocated for transmission probability is equivalent to the probability the BS is scheduled. Assume that the probability that a BS is scheduled is given such that a given throughput target is achieved. The proposed strategy then proceeds as follows:

1. Obtain the probability that BS k is scheduled, P_k, from throughput targets.

2. Obtain the scheduling weight for each BS by numerically solving the simultaneous nonlinear equations arising from Lemma 3.1.

3. Obtain the transmission time for each BS to meets its PLP target. Denote the set of transmission times $\{T_k\}_{k=1}^K$.

4. Obtain packet transmission time T by selecting the minimum transmission time such that $T = \min_k \{T_k\}_{k=1}^K$.

Observe that the design strategy provides a method to meet both throughput and PLP targets. Unfortunately, a simple closed-form expression for the PLP is only available for buffer size $B = 1$. For $B > 1$ semi-numerical techniques must be used to obtain an exact expression of the PLP [Egan et al., 2013b]. Another simple condition analogous to the PLP in the case of infinite buffer queues is queue stability. In the sequel, the difference is determined between the transmission time for the large buffer size regime and for the small buffer size regime. In the large buffer size regime the queue is required to be stable, and in the small buffer size regime the PLP target is required to be met.

3.4.2 Queue Stability and Packet Loss Probability

Queue stability is defined via the overflow function [Neely et al., 2002]

$$g(M) = \lim_{t\to\infty} \sup \frac{1}{t} \int_0^t 1_{[U(\tau)>M]} d\tau, \tag{3.16}$$

where $U(\tau)$ is the number of unprocessed packets in the queue at time τ. A queue is said to be stable if $g(M) \to 0$ as $M \to \infty$. Intuitively, this means that the queue does not grow unboundedly.

The following proposition provides a sufficient condition for queue stability.

Proposition 3.2. *Suppose that the probability of transmission for user k is P_k, the transmission time is T and the arrival rate is λ. A sufficient condition for queue stability is $\lambda_k < P_k/T$.*

Proof: The proof follows from [Neely et al., 2002] and the fact that the average service rate for BS k is P_k/T. ■

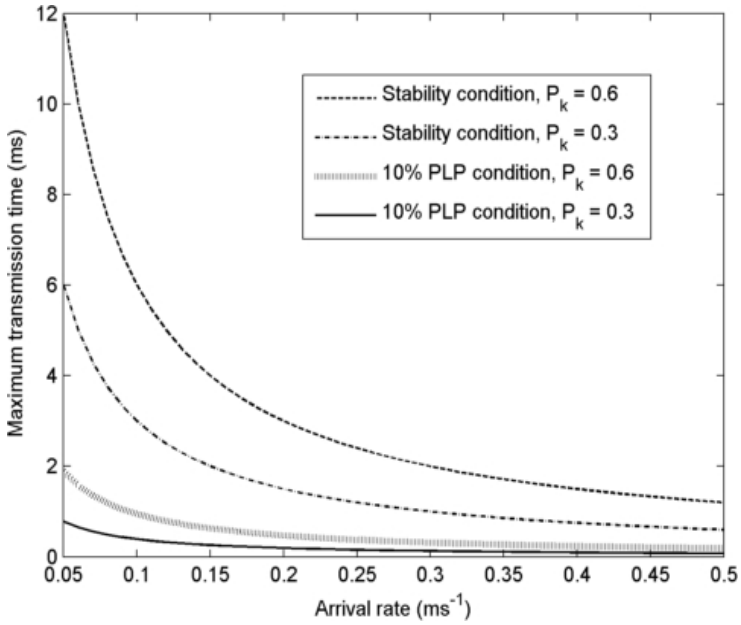


Figure 3.2: Plot of the maximum transmission time versus the arrival rate for varying probability of transmission. The transmission time design is based on: 1) queue stability using Proposition 3.2, and 2) a 10% PLP target using Proposition 3.1.

Note that this condition can be used to determine cases when the PLP approaches zero as the buffer size $B \to \infty$.

The transmission time required to guarantee queue stability (Proposition 3.2) is now compared with the transmission time required to guarantee the PLP bound is satisfied (Proposition 3.1). These form upper and lower bounds on the maximum transmission time, respectively. Fig. 3.2 plots the maximum transmission time that ensures the queue condition of queue k is satisfied for varying arrival rates and normalized service rates $P_k = 0.3$ and $P_k = 0.6$. To compute the transmission time corresponding to the PLP bound, the PLP target is $P_{L,k} \leq 0.1$. Observe that the maximum transmission time to guarantee queue stability is over 5 times larger than to guarantee the PLP target is met, for $\lambda_k = 0.05$ ms^{-1} and both normalized service rates. As such, queue stability should not be used for transmission time design in the small buffer regime.

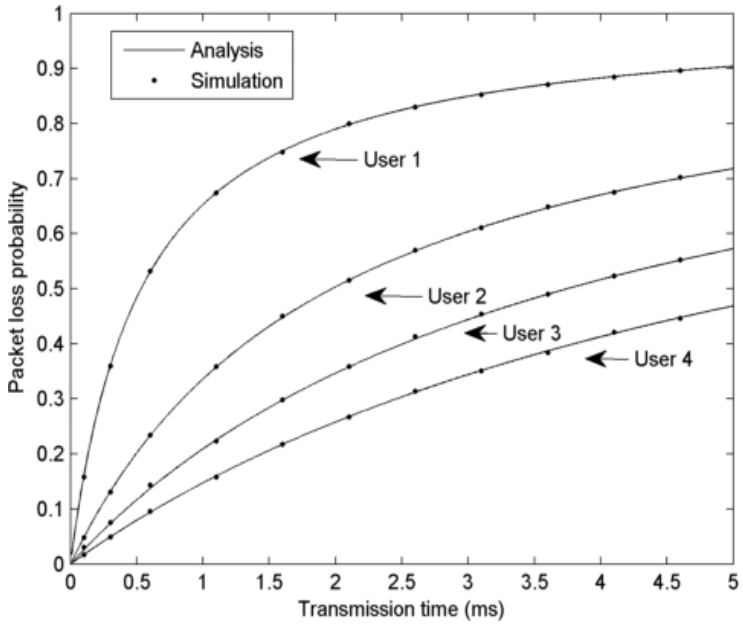


Figure 3.3: Plot of packet loss probability for each user versus the transmission time for $K = 4$ users with arrival rates $\lambda_i = 0.1$ ms^{-1}, $i = 1, 2, 3, 4$ and scheduling policy weights $W_1 = 1$, $W_2 = 2$, $W_3 = 3$, $W_4 = 4$.

3.5 Results and Discussion

The PLP analysis is confirmed via Monte Carlo simulation. Simulations are performed with $K = 4$ BSs, $B = 1$, packet arrival rates $\lambda_i = 0.1$ ms^{-1}, $1 \leq i \leq 4$ and scheduling weights $W_1 = 1$, $W_2 = 2$, $W_3 = 4$, $W_4 = 4$. Fig. 3.3 plots the PLP for each BS versus the transmission time. Note that the simulation curves are consistent with analysis. Observe that BS 1 has a larger PLP than the other BSs as $W_1 < W_i$, $i \neq 1$. Moreover, the PLP is observed to be increasing monotonically with the transmission time.

Fig. 3.4 plots the τ-achievable regions for fixed weight policies (corresponding to the τ_{FW}-achievable region) and exponential weight policies for $K = 2$ BSs. The exponential weight policy for BS k is

$$W_k = e^{c_k s_k}, \tag{3.17}$$

where s_k is the head of line (HOL) delay of BS k's packets and c_k is a constant coefficient for BS k that determines the probability BS k is scheduled. The arrival rates for each

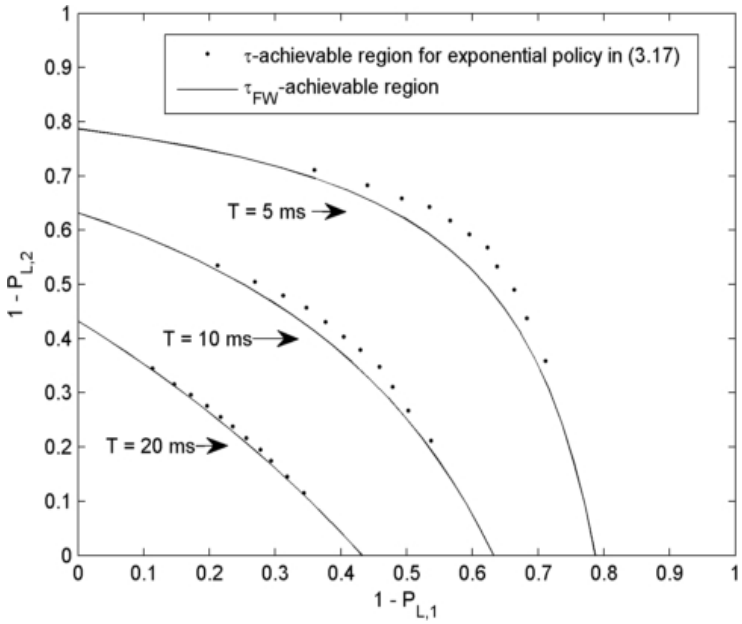


Figure 3.4: Plot of the τ_{FW}-achievable region and the τ-achievable region using the exponential policy described in (3.17) for $K = 2$ BSs with packet arrival rates $\lambda_1 = \lambda_2 = 0.1$ ms^{-1}

BS are $\lambda_1 = \lambda_2 = 0.1$ ms^{-1}. Observe that the τ-achievable regions for the exponential weight policy (obtained by varying c_k) are consistently larger than the τ_{FW}-achievable regions. This is expected, since the exponential weight policy accounts for the packet delay. The difference between the τ-achievable regions of the policies decreases as τ increases. Moreover, the difference between the τ-achievable regions decreases as one BS is prioritized. These observations suggest that the fixed weight policy can be implemented without PLP performance loss in networks with a long transmission time or with one BS that is heavily prioritized.

3.6 Conclusion

In this chapter, a fixed weight scheduling policy was proposed for BS scheduling in CoMP. A new simple closed-form bound was derived for the PLP for each user. A sufficient condition was then derived for queue stability for the network. A design strategy was developed for the scheduler weights and the transmission time, based on the PLP

bound. Finally, the τ-achievable region of the fixed weight policy was compared with an adaptive weight scheduling policy. Surprisingly, analysis and simulation showed that the proposed fixed weight scheduler can reduce communication overheads between BSs compared with the adaptive weight scheduler, while achieving a comparable PLP for each user.

Chapter 4

Structured Limited Feedback Codebook Design for Multi-user MIMO

A key component of multiuser MIMO using zero-forcing precoding is the feedback of quantized channel state information to the base station. A problem arises when each user has a common codebook as the quantized channels can form a singular matrix that results in a reduced sum-rate. In this chapter, two new structured constructions are proposed to generate different codebooks at each user via transformations of a base codebook. The first construction is based on the Householder transform, while the second proposed construction is based on the representation theory of groups.

4.1 Introduction

MU-MIMO an important technique to support high data rate downlink applications in multiuser wireless networks. This is due to the potential linear scaling in the achievable rate up to the order of the number antennas at the BS [Caire and Shamai, 2003]. In practice, the low complexity zero forcing (ZF) precoder is often used to suppress inter-user interference at the cost of a sub-optimal rate [Peel et al., 2005, Jindal, 2006]. To remove the inter-user interference perfectly, full channel state information (CSI) is required. Unfortunately, only imperfect CSI is available at the BS due to the limited rate CSI feedback channels commonly used in frequency division duplex (FDD) systems [Jindal, 2006].

Codebook-based quantization is a common technique used to feedback CSI over the low-rate feedback channel in the FDD MU-MIMO downlink [Jindal, 2006, Heath et al., 2009]. This feedback method is currently adopted in 3GPP LTE [Liu et al., 2012]. In

recent work, it has also been applied to network MIMO, where multiple cells cooperate via backhaul links [Liu et al., 2011, Bhagavatula and Heath, 2011, Zhou et al., 2011]. Unfortunately, when codebook-based quantization is combined with ZF precoding, the quantization error results in inter-user interference and sum-rate degradation. This is a significant problem in 802.16m and LTE as the quantization codebooks are typically limited to 2 or 4 bits [Li et al., 2010], which are known to perform poorly compared to the perfect CSI scenario [Heath et al., 2009].

A key problem for the downlink with low rate feedback, in addition to the quantization loss, occurs when different users select the same codeword to feedback. This is a particular problem when all users employ the same codebook, especially when the codebooks are small as in LTE. If more than one user selects the same codeword the channel matrix is rank deficient, and hence the sum-rate is reduced when ZF precoding is used. This is due to fact that one or more users is dropped, which has two consequences: a reduction of sum-rate; and an increase in delay for the dropped user, which is not tolerated in networks that support real-time services. Both consequences can be avoided by ensuring that each user employs a different codebook.

To avoid the rank deficiency problem, unique codebooks can be generated at each user by applying transformations to the common base codebook. This means that a different transformation is required at each user, in addition to the base codebook. To preserve the minimum mean square quantization error of the transformed codebook, the transformations must be unitary [Ding et al., 2007]. Approaches using random unitary transformations and the Householder transform were proposed in [Ding et al., 2007, Li et al., 2005]. The problem of constructing a number of different codebooks was also considered in [Jiang et al., 2012], where the random construction was employed. These approaches require either a different unstructured unitary matrix or unit-norm vector, for each user. Unfortunately, these approaches are inefficient as each user must store all different codebooks, since the appropriate codebook is not known when the user enters the cell. For example, using the random construction in [Ding et al., 2007] for a four antenna BS requires 154 kilobytes of memory at each user, which is significant – it is approximately the size of a software application on a smartphone.

In this chapter, two new structured unitary matrix constructions are proposed to generate the multiple different unitary transformations required to ensure a full-rank quantized channel matrix. The proposed structured constructions reduce the number of required unitary matrices stored in memory, for both the BS and each user. In fact, both the constructions only require a fixed number of transformations to be stored, independent of the number of users. In particular, the constructions require at most 1.1 kilobytes to store the codebooks at each user for a four antenna BS, which is less than 0.6% of the storage required for the random scheme in [Ding et al., 2007].

The first proposed construction exploits the Householder transform to ensure that each user has a unique codebook by transforming a common base codebook. Using the Householder transform is desirable as it is parameterized by only two parameters; both of which are already stored in memory. This means that the first construction requires no storage memory in addition to the base codebook. This is in contrast to the approach in [Li et al., 2005], which also uses the Householder transform but relies on storing a large number of vectors at each user. A new condition is also derived that ensures the elements in the base codebook and each transformed codebook are unique.

As an additional benefit, the proposed Householder construction is shown to reduce the search complexity when the well-known Fourier base codebook [Love et al., 2003] is used. In particular, the number of codebook search multiplications and additions are shown to be reduced by up to 50% using the first construction, compared with the standard Fourier base codebook [Love et al., 2003].

The second construction exploits unitary group representations (GR) of *fixed point free groups* [Shokrollahi et al., 2001]. This construction is necessary when the base codebook has a group structure, such as the Fourier base codebook [Love et al., 2003], for which the Householder does not provide multiple unique transforms. The GR construction has low storage requirements as no more than four matrices must be stored, with each matrix element drawn from a finite field.

To compare the constructions with the standard case of using a common codebook at each user, a lower bound on the rate outage probability is derived for the common codebook scheme. The lower bound on the rate-outage probability is shown to increase

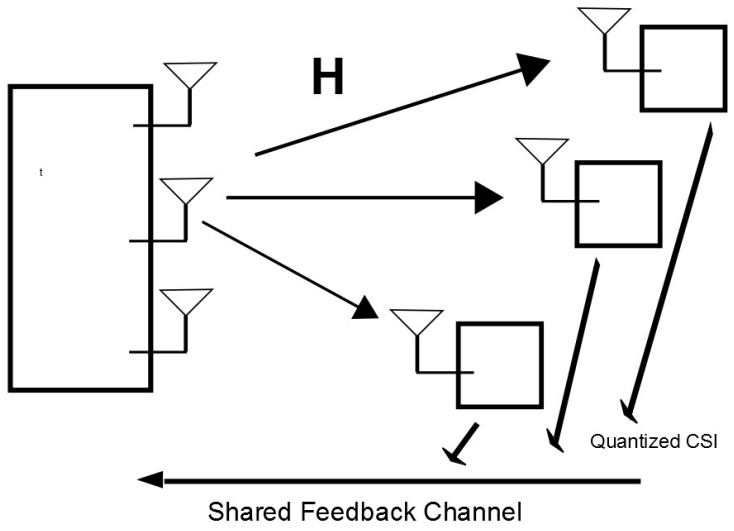


Figure 4.1: MU-MIMO downlink system model.

as the codebook size is reduced, indicating that the constructions are necessary in the small-codebook regime. The rate-outage probability bound is also shown to be minimized when every codeword is equally likely, for a fixed quantization codebook size. Many equiangular tight frames (ETFs) are proven to satisfy this property, and are later used in the form of the optimal ETF base codebook in simulations of the GR construction, the random construction and the common codebook scheme.

Finally, Monte Carlo simulations are performed to verify the performance of the proposed constructions. Both of the proposed constructions are shown perform comparably with the intuitive but impractical random transformation construction in [Ding et al., 2007], and do so with significantly reduced storage requirements. Most importantly, the figures show that the schemes provide a significant improvement over common codebooks at each user, for practical codebook sizes. For example, when 4 bits of feedback are used, the Householder construction outperforms the common codebook scheme by 25% when the BS employs 4 antennas.

4.2 System Model

Consider the MU-MIMO downlink (see Fig. 4.1, where a BS with N_t antennas simultaneously transmits to $K \leq N_t$ users with a single antenna each. The received signal vector is denoted by $\mathbf{y} = [y_1, \ldots, y_K]^T$, given by

$$\mathbf{y} = \sqrt{P}\mathbf{H}\mathbf{V}\mathbf{s} + \mathbf{n}, \tag{4.1}$$

where $\mathbf{V} = \sqrt{P}[\mathbf{v}_1, \ldots, \mathbf{v}_{N_t}]$ is the precoder, P is the transmit power, $\mathbf{H}^\dagger = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$ is the channel matrix for which $\mathbf{h}_i^\dagger$ corresponds to the channel vector of the i-th user, and $\mathbf{s}$ denotes the Gaussian data symbols, $\mathbf{n}$ is the noise vector with each element distributed as $n_i \sim \mathcal{CN}(0, \sigma^2)$, and σ^2 is the noise power at each user. The inverse of the noise power is defined as $\rho = 1/\sigma^2$, which is closely related to the per-user SNR.

The signal received by user i is

$$y_i = \sqrt{P}\mathbf{h}_i^\dagger \mathbf{v}_i s_i + \sum_{k=1, k\neq i}^{N_t} \sqrt{P}\mathbf{h}_i^\dagger \mathbf{v}_k s_k + n_i, \tag{4.2}$$

where $\mathbf{h}_i^\dagger$ is the i-th row of the channel matrix $\mathbf{H}$, known as the channel vector for the i-th user.

In this chapter, the ZF precoder is used to suppress interuser interference. Assume that a perfect estimate of the corresponding channel vector is obtained at each user and fed back to the BS to derive the ZF precoder. Each user's CSI is assumed to be fed back over a low-rate, error free, zero-delay feedback channel. The quantized channel matrix $\hat{\mathbf{H}}$ is derived from each user's quantized channel vector. The ZF precoder with equal power allocation is then by

$$\mathbf{V} = \frac{\hat{\mathbf{H}}^\dagger(\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger)^{-1}}{\|\hat{\mathbf{H}}^\dagger(\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger)^{-1}\|_F}, \tag{4.3}$$

where $\|\cdot\|_F$ is the Frobenius matrix norm. Note that the rate reduction using equal power allocation is negligible compared with the optimal waterfilling scheme at high signal-to-noise ratio (SNR), for a sufficiently large codebook size [Peel et al., 2005]. We note that the precoder $\mathbf{V}$ does not guarantee that each user receives the same transmit power. Instead, the transmit power for each user is dependent on $\mathbf{H}$, while ensuring that the total transmit

power constraint is satisfies. This approach follows [Peel et al., 2005].

To quantize the channel vectors, user i decomposes $\mathbf{h}_i^\dagger$ into the channel gain $\|\mathbf{h}_i^\dagger\|$ and the channel shape

$$\tilde{\mathbf{h}}_i^\dagger = \frac{\mathbf{h}_i^\dagger}{\|\mathbf{h}_i^\dagger\|}. \tag{4.4}$$

To quantize the channel shape, user i searches through the codebook $\mathcal{F}_i = \{\mathbf{f}_{i,k}\}_{k=1}^N$ spanning $\mathbb{C}^{N_t}$, where N is the number of codewords in the codebook and $\mathbf{f}_{i,k}$ are the codewords. The index of the chosen codeword for user i, k_i^*, maximizes the inner product search criterion as in [Jindal, 2006],

$$k_i^* = \arg \max_{k=1,\ldots,N} |\tilde{\mathbf{h}}_i^\dagger \mathbf{f}_{i,k}|^2,\ i = 1, \ldots, K. \tag{4.5}$$

Throughout this chapter, different codebooks are constructed for each user by applying unitary transformations to the codebook for the user with index $i = 1$. More specifically, codebooks are constructed with codewords that are unique in $\mathbb{C}P^{N_t-1}$, which represents the complex projective space and is equivalent to the Grassmannian manifold $\mathcal{G}(N_t, 1)$. The codebook of user $i = 1$ is referred to as the *base* codebook. The codebooks for the other users, obtained by applying unitary transformations to the base codebook, are referred to as *transformed* codebooks. All users' codebooks are written in terms of the base codebook by denoting the base codebook as $\mathcal{F}_1 = \mathcal{F}$, and the codewords for the codebook of user i as

$$\mathbf{f}_{i,k} = \mathbf{U}_i \mathbf{f}_k,\ k = 1, \ldots, N,\ i = 1, \ldots, K, \tag{4.6}$$

where $\mathbf{U}_i$ is the i-th users unitary transformation satisfying $\mathbf{U}_i \mathbf{U}_i^\dagger = \mathbf{U}_i^\dagger \mathbf{U}_i = \mathbf{I}$. In the special case of a common codebook for each user $\mathbf{U}_i = \mathbf{I},\ i = 1, \ldots, K$.

An important base codebook is the Fourier codebook. The Fourier codebook can be easily constructed using the direct sum to obtain a reducible representation of a cyclic group generated by

$$\mathbf{U} = \chi_1 \oplus \cdots \oplus \chi_d$$

$$= \begin{pmatrix} e^{2\pi j k_1/N} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & e^{2\pi k_d/N} \end{pmatrix}, \tag{4.7}$$

where $\oplus$ is the direct sum and $\chi_i = e^{2\pi j k_i/N}$ for some $1 \leq k_i \leq N-1$. Hence, each codeword in the Fourier codebook is given by $\mathbf{f}_l = \mathbf{U}^l \mathbf{Œ}$, $l = 0, \ldots, N-1$, where $\mathbf{Œ} = 1/\sqrt{d}[1, 1, \ldots, 1]^T$.

The performance of the MU-MIMO downlink with limited feedback is evaluated using the sum-rate, which is given by [Peel et al., 2005]

$$R = \sum_{i=1}^{K} \log_2 \left(1 + \frac{P\rho |\mathbf{h}_i^\dagger \mathbf{v}_i|^2}{P\rho \sum_{k=1, k\neq i}^{N_t} |\mathbf{h}_i^\dagger \mathbf{v}_k|^2 + 1} \right). \tag{4.8}$$

Sum-rate outage occurs when $R < R_0$, for a sum-rate threshold R_0. Similarly, a per-user outage event is defined, which occurs when

$$R_i = \log_2 \left(1 + \frac{P\rho |\mathbf{h}_i^\dagger \mathbf{v}_i|^2}{P\rho \sum_{k=1, k\neq i}^{N_t} |\mathbf{h}_i^\dagger \mathbf{v}_k|^2 + 1} \right) < R_{0,i}, \tag{4.9}$$

where $R_{0,i}$ is the per-user rate threshold for user i.

4.3 Proposed Codebook Constructions

In this section, two new constructions are proposed to obtain different codebooks for each user: the Householder construction; and the group representation construction. The Householder construction employs the Householder transform and does not require any storage memory, in addition to the base codebook. The Householder construction is generally applicable to a wide range of base codebooks, including those generated using the Lloyd algorithm [Xia et al., 2005]. Unfortunately, in the special case that the base codebook is the Fourier base codebook, the Householder construction does not yield different codebooks at each user. To overcome this, the new GR construction is proposed to obtain the different codebooks at each user. Later, the storage requirements of the two constructions are analyzed. Finally, applying the Householder construction to the Fourier base

codebook is shown to be beneficial, despite only yielding two different codebooks. In particular, the search complexity is shown to be reduced by up to 50% compared with the standard Fourier base codebook.

4.3.1 Householder Construction

The Householder construction employs the complex Householder transform. This is desirable as the Householder transform requires only two vectors: an initial vector and a transformed vector. The initial vector is chosen to be a codeword from the base codebook and the transformed vector is chosen to be $\mathbf{e}_1 = [1, 0, 0, \ldots]^\dagger$. This means that both the initial vector and the transformed vector can be obtained either directly from the base codebook or a standard vector. As such, no additional memory space is required to construct different codebooks at each user, in addition to the base codebook.

The Householder construction, starts with a base codebook. Possible choices are the random construction [Jindal, 2006], the Lloyd algorithm [Xia et al., 2005], or the optimized base codebook detailed in Section 4.4. The complex Householder transform is then applied and is given by [Kuo-Liang and Wen-Ming, 1997]

$$\mathbf{U}_H(\mathbf{a}) = \mathbf{I} - \frac{\mathbf{z}\mathbf{z}^\dagger}{\mathbf{z}^\dagger\mathbf{a}}, \tag{4.10}$$

where $\mathbf{a}$ is the initial vector that is an element of the base codebook, $\mathbf{b} = \mathbf{e}_1$ is the transformed vector and $\mathbf{z} = \mathbf{a} - \mathbf{b}$. The Householder transform is both unitary and Hermitian when $\|\mathbf{a}\| = \|\mathbf{b}\|$, as is the case here. The third step in the Householder construction, is to allocate each user a different transformed codebook. This is achieved for user i by taking codeword $\mathbf{f}_i$ from the base codebook, and computing the unitary transformation matrix in (4.10) with the initial vector chosen to be $\mathbf{a} = \mathbf{f}_i$. The transformed codebook for user i is then constructed by multiplying the unitary matrix computed using $\mathbf{f}_i$ with every codeword in the base codebook. In short, the final form of transformed codebooks constructed using the Householder construction is given by $\mathbf{U}_H(\mathbf{f}_i)\mathbf{F}$ for user i, with base codebook $\mathbf{F}$. Each step is further detailed in Table 4.1.

A key feature of the Householder construction is that no additional storage is re-

Table 4.1: Householder Construction Quantization Algorithm

1) Choose a base codebook. This could be random, generated using the Lloyd algorithm [Xia et al., 2005].

2) The BS allocates the each user an index $k \in \{1, 2, \ldots, N\}$.

3) The user k performs perfect channel estimation to obtain channel shape $\tilde{\mathbf{h}}$.

4) The user k constructs the Householder transformation $\mathbf{U}_H(\mathbf{a})$ from (4.10) with $\mathbf{a} = \hat{\mathbf{h}}_k$ and $\mathbf{b} = [1, 0, \ldots, 0]^T$. The transformed codebook of user k is given by $\mathbf{U}_H(\mathbf{f}_k)\mathbf{F}$.

5) The user k performs channel quantization on the channel shape $\tilde{\mathbf{h}}$ as in (4.5). I.e. selects the codeword $i^* = \arg\max_{i=1,2,\ldots,N} |\mathbf{f}_i^\dagger \mathbf{U}_H(\mathbf{f}_k)^\dagger \tilde{\mathbf{h}}|$.

quired, in addition to the base codebook. This is the case because the transformation requires only vectors from the base codebook and these vectors are already stored at each user and the BS.

The Householder construction yields unique transformed codebooks for a wide range of base codebooks including those generated using the Lloyd algorithm [Xia et al., 2005] or the random approach [Jindal, 2006]. To see this, a transformed codebook using the Householder construction is unique if $\mathbf{U}_H(\mathbf{f}_h)\mathbf{f}_k \neq \mathbf{U}_H(\mathbf{f}_i)\mathbf{f}_l e^{j\theta}$ for a $\theta \in [0, 2\pi)$, all users $h, i = 1, 2, \ldots, K,\ h \neq i$, and all codewords $\mathbf{f}_k,\ \mathbf{f}_l,\ k, l = 1, 2, \ldots, N$. There are two situations that the uniqueness requirement is violated. First, if the condition for uniqueness $\mathbf{f}_k \neq \mathbf{f}_l e^{j\theta}$ is not satisfied for a $\theta \in [0, 2\pi)$ and a pair (k, l), $k, l \in \{1, 2, \ldots, N\}$. This condition for uniqueness is satisfied for practical constructions such as the Lloyd algorithm and is satisfied with probability 1 for the random base codebook.

Before the second situation is described where the uniqueness requirement is violated, note that there always exists a unitary matrix $\mathbf{U}_{ji}$ such that $\mathbf{U}_{ji}\mathbf{f}_j = \mathbf{f}_i$ for codewords i, j. As a consequence, the Householder transforms in the construction can be written in terms of one codeword. In particular, define $\mathbf{U}_H(\mathbf{f}_1)$ such that $\mathbf{U}_H(\mathbf{f}_1)\mathbf{f}_1 = \mathbf{e}_1$, where $\mathbf{U}_H(\mathbf{f}_1)$ is the transformation matrix in the Householder construction for user 1. Since $\mathbf{U}_{ji}\mathbf{f}_j = \mathbf{f}_i$, the transformation matrix for user i is obtained as

$$\mathbf{U}_H(\mathbf{f}_i) = \mathbf{U}_H(\mathbf{f}_1)\mathbf{U}_{i1} = \mathbf{U}_H(\mathbf{f}_1)\mathbf{U}_{1i}^{\dagger}. \tag{4.11}$$

Now, the second situation where the uniqueness requirement is violated occurs when there are codewords in the base codebook $\mathbf{f}_k, \mathbf{f}_l$ that violate the condition that $\mathbf{U}_{1i}^{\dagger}\mathbf{U}_{1k} \neq \mathbf{U}_{1m}^{\dagger}\mathbf{U}_{1l}e^{j\theta}$ for fixed i, m, a $\theta \in [0, 2\pi)$ and a pair (k, l), $k, l \in \{1, 2, \ldots, N\}$, $k \neq l$. This follows directly from (4.11) by substituting in the transformation matrices arising for different users into (4.10). Fortunately, this condition on uniqueness is satisfied for the Lloyd algorithm and for the random base codebook.

There is one scenario where the uniqueness conditions are violated for the Householder construction. The scenario occurs when the unitary matrices $S = \cup_{\theta \in [0, 2\pi)} \{e^{j\theta}\mathbf{U}_{1i}\}_{i=1}^{N}$ form a group of unitary matrices. That is, the matrices along with the matrix multiplication operation satisfy closure, associativity, and existence of inverse and identity proper-

ties. In this case, there is always a matrix $\mathbf{U}_{1n} \in \{\mathbf{U}_{1i}\}_{i=1}^{N}$ that satisfies $\mathbf{U}_{1n} = e^{j\theta}\mathbf{U}_{1i}^{\dagger}\mathbf{f}_{1k}$ for a $\theta \in [0, 2\pi)$. As a consequence, the second condition for uniqueness is violated, which means that there is at least one codeword in user i's codebook that is in the same subspace as a codeword in user j's codebook for a pair (i, j), $i, j \in \{1, 2, \ldots, N\}$, $i \neq j$.

The uniqueness condition violation is entirely dependent on the base codebook. In particular, the violation is guaranteed to occur when the base codebook is constructed using a group of unitary matrices $\{\mathbf{U}_i\}_{i=1}^{N}$ and multiplying each group element with an initial vector $\mathbf{f}_1$. This type of base codebook with group structure is explicitly written as $\mathbf{F} = [\mathbf{U}_1\mathbf{f}_1, \ldots, \mathbf{U}_N\mathbf{f}_1]$. A common example of a base codebook with group structure is the Fourier base codebook [Love et al., 2003]. In order to ensure that the Householder construction yields more than two unique codebooks, base codebooks with group structure cannot be used.

4.3.2 Group Representation Construction

The new GR construction is proposed to construct unique transformed codebooks when the base codebook has group structure, for which the Householder construction does not yield unique codebooks. Although the GR construction is also applicable for other types of base codebooks, it is preferable to employ the Householder construction for these base codebooks as the GR construction has larger storage requirements. The construction follows that of the Householder construction, but instead of Step 4) of Table 4.2 unitary representations of groups are exploited to derive unique codebooks for each user.

Three definitions are required from [James and Liebeck, 2001] to detail the GR construction and prove that the transformed codebooks are unique. In the following, the abstract formulation of a group is a set G together with a binary operation that satisfies closure, associativity, and existence of inverse and identity properties. Concrete group constructions are provided later in Example 1 and Example 2.

In order to define a unitary representation–the building block of the construction–an important mapping between a groups, known as a homomorphism, is required.

Definition 4.1. *Let G and F be groups. Then, a homomorphism from G to F is a function*

$\theta : G \to F$ which satisfies

$$\theta(g_1 g_2) = \theta(g_1)\theta(g_2), \text{ for all } g_1, g_2 \in G. \tag{4.12}$$

Unitary representations take an element from the abstract group G and map it to a unitary matrix such that the map is a homomorphism. This means that the identity of G is mapped to the identity of the representation and inverses of elements in G are mapped to inverses of elements of the representation, ensuring that the group structure is preserved. Formally, a representation is defined as follows.

Definition 4.2. *Let G be a group. A representation of G over $\mathbb{C}$ is a homomorphism ρ from G to the general linear group $GL(n, \mathbb{C})$ (the group of $n \times n$ matrices over the complex field) [James and Liebeck, 2001], for some integer n. A representation is faithful if the identity element of G is the only element for which $\theta(g) = \mathbf{I}$, $g \in G$.*

The GR construction utilizes group representations to obtain unitary matrices that construct different codebooks at each user. To ensure that the codebooks are unique for arbitrary base codebooks, the group representations are required to be *fixed point free* groups. A fixed point free group is defined as follows [Shokrollahi et al., 2001].

Definition 4.3. *For some $N_t \in \mathbb{N}$, a fixed point free group G has an N_t-dimensional representation that satisfies*

$$\mathbf{U}_i \mathbf{x} \neq \mathbf{x}, \ \forall \mathbf{U}_i \in G, \ \mathbf{x} \in \mathbb{C}^{N_t}. \tag{4.13}$$

The condition in (4.3) is equivalent to requiring that any representation of the group G does not have elements with an eigenvalue $\lambda = 1$, as otherwise there would exist an eigenvector, $\mathbf{x}$, of a $\mathbf{U}_i \in G$ such that $\mathbf{U}_i \mathbf{x} = \mathbf{x}$.

The fixed point free property is now shown to be a necessary condition for the uniqueness of the transformed codebooks.

Proposition 4.1. *Let the base codebook be an arbitrary set of unit norm vectors in $\mathbb{C}^{N_t}$ and $G = \{\mathbf{U}_i\}$ be a group of unitary matrices. A necessary condition for the transformed codebook to*

be unique is that there exists no codewords such that $\mathbf{U}_i\mathbf{f}_k = \mathbf{f}_l$*, where* $\mathbf{U}_i \in G$*, which is satisfied when* G *is fixed point free.*

Proof. In order for the transformed codebook to be unique in $\mathbb{C}P^{N_t-1}$, no codewords can exist such that

$$\mathbf{U}_i\mathbf{f}_k = \mathbf{f}_l, \tag{4.14}$$

for all $\mathbf{f}_k$, $\mathbf{f}_l$ in the base codebook and $\mathbf{U}_i \in G$. Suppose that the representation $\{U_i\}$ is not faithful. Then, there exists $\mathbf{U}_1 = \mathbf{U}_j = \mathbf{I}$ for some $j \in \{1, \ldots, |G|\}$. As such, the transformed codebooks corresponding to $\mathbf{U}_j$ is the same as the base codebook, which means that it is not unique. Now suppose that G is not fixed point free. This means that there exists a $\mathbf{x} \in \mathbb{C}^{N_t}$ such that $\mathbf{U}_i\mathbf{x} = \mathbf{x}$ for some $\mathbf{U}_i \in G$. If $\mathbf{x}$ is in the base codebook, then the transformed codebook is not unique. As a consequence, a necessary condition for the transformed codebook to be unique for all base codebooks is satisfied when the group representation $\{U_i\}$ of G is a faithful homomorphism and G is fixed point free. □

Proposition 4.1 provides a necessary condition for the GR construction to generate different codebooks at each user, for an arbitrary base codebook. A full classification of representations of fixed point free groups is given in [Shokrollahi et al., 2001].

To ensure that the codebooks are unique it is required that no codewords lie in the same subspace. Although the fixed point free property is a necessary condition for uniqueness, it does not by itself guarantee that the group elements satisfy

$$\mathbf{U}_i\mathbf{v} \neq e^{j\theta}\mathbf{U}_j\mathbf{v}, \tag{4.15}$$

for all $\mathbf{U}_i, \mathbf{U}_j \in G$ $i \neq j$. In fact, it only guarantees that $\mathbf{U}_i\mathbf{v} \neq \mathbf{U}_j\mathbf{v}$ for all $\mathbf{v} \in \mathbb{C}^{N_t}$.

Fortunately, this problem can be easily overcome by systematically removing transformations. Although this reduces the number of possible transformed codewords, many fixed point free groups are extremely large, with over 1000 elements. As such, this does not affect the number of users that can be allocated to the base station.

Now, the systematic removal of transformations is detailed to ensure (4.15) is satis-

fied. First, find elements of the representation with the form $e^{j\theta}\mathbf{I}$. The set of such elements forms a normal subgroup N, which defines an equivalence relation. Each set of equivalent elements in G is given by the cosets $gN,\ g \in G$. To obtain the final set of unique transformations, a representative element of each coset is chosen (any coset element can be chosen). This procedure ensures that the GR construction has different codebooks at each user even if the base codebook is the Fourier codebook.

The described systematic removal method is illustrated in the following concrete examples of fixed point free group representations. The first example of a fixed point free group is the cyclic group. An important property of the GR construction based on the cyclic group is that it allows for the scheme to be applied to the MU-MIMO downlink with an arbitrarily large number of users. In particular, a fixed point free unitary representation of a cyclic group exists for all $N \geq 1$.

Example 4.1. *Cyclic groups have a presentation $C_n = \langle a|a^n = 1\rangle$. As a result, the unitary representation can be obtained simply using a single matrix*

$$\mathbf{A} = \text{diag}\left(\exp\left(2\pi j\frac{k_1}{n}\right), \ldots, \exp\left(2\pi j\frac{k_{N_t}}{n}\right)\right), \tag{4.16}$$

where $k_1, \ldots, k_{N_t} < n$ are distinct and each k_i is relatively prime to n. Furthermore, the k_i must all satisfy $k_i \neq k_1 + \frac{mn}{l'}$, for some $m = 1, 2, \ldots$ and a fixed $0 < l' < n$. This ensures that condition (4.15) is satisfied. The elements of the representation are then given by $\mathbf{U}_{G,i+1} = \mathbf{A}^i,\ i = 0, \ldots, n-1$. This transformation is applicable for $N_t > 1$.

Note that care is required when using the cylic group construction as the Fourier codebook [Love et al., 2003] also has the same structure. This means that the coefficients $k_i,\ i = 1, \ldots, N_t$ must be chosen such that no group elements are the same as for the base codebook.

Another class of fixed point free group (as in Definition 4.3) is $G_{m,r}$ that is explicitly defined in [Shokrollahi et al., 2001]. In the second example, $G_{6,-1}$ is used, which has a 2-dimensional representation.

Example 4.2. *The group $G_{6,-1}$ has a representation $\mathbf{A}^k\mathbf{B}^l$, $k = 0, \ldots, 5$, $l = 0, 1$, where*

$$\mathbf{A} = \begin{pmatrix} \zeta & 0 \\ 0 & \zeta^{-1} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{4.17}$$

where $\zeta = e^{j2\pi/6}$. Note that $G_{6,-1}$ has a normal subgroup $N = \{\mathbf{I}, e^{j\pi}\mathbf{I}\}$, where $\mathbf{A}^3 = e^{j\pi}\mathbf{I}$. The normal subgroup satisfies $gN = Ng$, $g \in G_{6,-1}$, which is equivalent to requiring left and right cosets of N in G are equivalent. Now define the equivalence relation $\mathbf{A} \sim e^{j\theta}\mathbf{A}$. The cosets of the normal subgroup are then equivalence classes. In particular,

$$\begin{aligned} &E_1 = N,\ E_2 = \{\mathbf{A}, \mathbf{A}^4\},\ E_3 = \{\mathbf{A}^2, \mathbf{A}^5\} \\ &E_4 = \{\mathbf{B}, \mathbf{A}^3\mathbf{B}\},\ E_5 = \{\mathbf{AB}, \mathbf{A}^4\mathbf{B}\},\ E_6 = \{\mathbf{A}^2\mathbf{B}, \mathbf{A}^5\mathbf{B}\}. \end{aligned} \tag{4.18}$$

By choosing a representative element from each equivalence class, the condition (4.15) is satisfied. In particular, a set of valid transformations is then $\{\mathbf{I}, \mathbf{A}, \mathbf{A}^2, \mathbf{B}, \mathbf{AB}, \mathbf{A}^2\mathbf{B}\}$.

The GR construction has been shown to yield a large number of unique transformations, and now it is explained how codebook transformations are generated and allocated to users. First, choose a base codebook $\mathbf{F}$ and elements of the unitary representation of a fixed point free group (defined in Definition 4.3), denoted $\{\mathbf{U}_{G,k}\}_{k=1}^{N}$. The procedure to obtain $\{\mathbf{U}_{G,k}\}$ is detailed in Step 4 of Table 4.2. Then, the BS allocates each user an index $k \in \{1, 2, \ldots, |G|\}$. Each user k then constructs the corresponding unitary transformation, given by $\mathbf{U}_{G,k}\mathbf{F}$. The transformed codebooks are guaranteed to be unique since the transformations are chosen to satisfy the conditions in Proposition 4.1 and (4.15). Two potential unitary transformations are in Example 1 and Example 2. Each user k then quantizes the channel shape by choosing a codeword from its transformed codebook $\mathbf{U}_{G,k}\mathbf{F}$ as per Step 5) of Table 4.2.

Note that the new GR construction is applicable to all base codebooks as the GR construction satisfies the conditions in Proposition 4.1 and (4.15). Moreover, the storage requirements for the GR construction do not grow with the number of users, which is the case for the random [Ding et al., 2007] and "improved Hochwald" [Li et al., 2005] constructions.

Table 4.2: Group Representation Construction Quantization Algorithm

1) Choose a base codebook. This could be random, generated using the Lloyd algorithm [Xia et al., 2005], or generated using the optimized codebook in Section 4.4.

2) The BS allocates each user an index $k \in \{1, 2, \ldots, |G|\}$.

3) User k performs perfect channel estimation to obtain channel shape $\tilde{\mathbf{h}}$.

4) User k constructs the transformed codebook using the unitary representation of the chosen group G (e.g. the fixed point free group in Example 1 or Example 2). To obtain the transformed codebook:

   a) Find elements in the representation of G of the form $e^{j\theta}\mathbf{I}$.

   b) Construct the normal subgroup N consisting of the elements of G obtained in Step 4a.

   c) Construct cosets gN, $g \in G$.

   d) Construct the final set of transformations $\{U_{G,k}\}$ by choosing a representative element from each coset.

   e) Construct the transformed codebook for user k using with base codebook $\mathbf{F}$, which is given by $U_{G,k}\mathbf{F}$

5) User k performs channel quantization on the channel shape $\tilde{\mathbf{h}}$, $k = 1, \ldots, |G|$ as in (4.5). I.e. the user selects the codeword $i^* = \arg\max_{i=1,2,\ldots,N} |\mathbf{f}_i^\dagger \mathbf{U}_{G,k}^\dagger \tilde{\mathbf{h}}|$.

4.3.3 Uniqueness Properties of the Proposed Constructions and Storage Requirements

In this section, the commonalities and differences of the constructions are detailed in terms of codebook uniqueness and storage requirements. In particular, the base codebooks are classified for which each of the proposed codebook constructions yield multiple unique transformed codebooks, conditions on the base codebooks to ensure that each transformed codebook is unique, and provide a comparison of the memory requirements of the proposed codebook construction with the random construction [Ding et al., 2007].

The main difference between the Householder and GR constructions is the type of base codebooks that yield unique transformed codebooks. As detailed earlier, the Householder construction is applicable to base codebooks that do not have group structure. An important example of a base codebook that has group structure is the Fourier codebook [Love et al., 2003]. The GR construction overcomes the group structure problem and can in fact be applied to the Fourier base codebook.

In general, it is not straightforward to check if a given base codebook has group structure without directly checking the codebook's Gram matrix. To this end, new conditions are developed that ensure that each transformed codebook is unique. The conditions can be used to check an arbitrary set of unitary transformations to see if the transformed codebooks are unique. As such, the conditions can be used to check if the Householder construction yields unique transformed codebooks. The following theorem explicitly provides the uniqueness conditions.

Theorem 4.1. *Let $\mathbf{U}_T$ be the Householder construction or the GR construction for any user, where the base codebook is $\mathcal{F} = \{\phi_i\}_{i=1}^N$. Let $\mathbf{U}_{ki}\phi_k = \phi_i,\ i,k = 1,\ldots,N$. The elements in the base codebook $\mathcal{F}$ and the elements in the transformed codebook $\{\mathbf{U}_T\phi_i\}_{i=1}^N$ for each user are unique if*

1. *$\phi_i \neq e^{j\theta}\phi_k,\ \forall k \neq i$.*

2. *No elements of the base codebook $\{\phi_i\}_{i=1}^N$ are eigenvectors of $\mathbf{U}_T\mathbf{U}_{ki},\ k = 1,\ldots,N$.*

Proof: See Appendix B.1. ■

Note that Theorem 4.1 holds for both the Householder and GR constructions. To illustrate the use of Theorem 4.1, consider the base codebook with structure $\mathbf{F} = [\mathbf{U}_1\mathbf{f}_1, \ldots, \mathbf{U}_N\mathbf{f}_1]$, where $\{\mathbf{U}_i\}$ forms a group of $N_t \times N_t$ unitary matrices and $\mathbf{f}_1$ is a unit norm N_t-dimensional vector. In this case the theorem can be applied by obtaining the eigenvectors of $U_T U_i$ (U_T is obtained using the Householder construction for some codework $\mathbf{U}_i\mathbf{f}_1$). Each eigenvector is then checked to ensure that it is not equivalent to any elements of the base codebook $\mathbf{F}$. Since the base codebook has group structure, there is in fact at least one codeword from that base codebook that is equivalent to the eigenvector, and hence the transformed codebooks are not unique.

A comparison is now performed of the storage requirements of the proposed codebook constructions. The GR construction only requires at most $64N_t^2$ bytes of storage, and often less due to the fact that the elements are often drawn from finite fields. This is due to the fact that there are 8 bytes in a double precision floating point number and every element is complex in the matrices corresponding to the GR construction. The factor 64 arises as there are at most four matrices used to construct the different codebooks for the GR construction. The Householder construction requires no storage in addition to the base codebook. In contrast, the storage requirement of the random construction is $16K_N N_t^2$ bytes, where K_N is the total number of users in the network, and N_t is the number of transmit antennas.

In LTE cells the number of users is typically at least 200 [Motorola, 2007]. Consider the scenario with 3 adjacent cells with 200 user each. In this case, the users in the network is $K_N = 600$ and the required storage requirement for the random construction with $N_t = 4$ is 154 kilobytes, which is significant – it is approximately the size of a software application. Furthermore, BSs with large antenna arrays are expected to play an important role in future cellular networks employing massive MIMO [Marzetta, 2010] and also rural wireless networks [Suzuki et al., 2012]. In these systems, BSs can have $N_t = 32$ or more antennas. In this case, the storage requirement for the random construction rises to 9.8 megabytes, using the same network parameters as in the previous LTE example. Certainly, this is prohibitive in practice.

4.3.4 Sparsity Properties of the Proposed Constructions

In this section, new sparsity properties are proven for the Householder construction in the case that the base codebook (used to generate the codewords for each user) is the Fourier codebook. The transformed codebooks are obtained using the final form of the Householder construction. Although the Householder construction applied to the Fourier codebook does not yield multiple unique codebooks, it is shown to be useful when only one common codebook is employed at each user. In particular, the key finding of the sparsity analysis is that the search complexity of the Fourier codebook is reduced by up to half compared with the standard approach in [Love et al., 2003].

To obtain the sparsity results, first the notion of a frame [Strohmer and Heath, 2003] and the relationship to quantization codebooks in MU-MIMO are described.

Definition 4.4. *A sequence $\mathcal{F} = \{\mathbf{f}_i\}_{i=1}^N$ is called a frame for the Hilbert space $\mathcal{H} = \mathbb{C}^d$ or $\mathbb{R}^d$, if there exists constants $0 < A \leq B < \infty$ such that $A\|\mathbf{x}\|^2 \leq \sum_{i=1}^N |\langle \mathbf{x}, \mathbf{f}_i \rangle|^2 \leq B\|\mathbf{x}\|^2$, for all $\mathbf{x} \in \mathcal{H}$.*

Observe that if the dimension of a sequence of vectors is finite, the sequence of vectors spans $\mathbb{C}^d$, and each vector satisfies $\|\mathbf{f}_i\|^2 = 1,\ i = 1, 2, \ldots, N$, then the sequence of vectors is a frame. Since the quantization codebook has $d = N_t < \infty$ and unit norm elements spanning $\mathbb{C}^d$, the codebook forms a frame.

Next, the sparsity properties of frames, and hence quantization codebooks, are quantified. A definition is required, based on [Casazza et al., 2011]. Note that the frame synthesis matrix of a codebook $\mathcal{F} = \{\mathbf{f}_i\}_{i=1}^N$ is given by $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N]$.

A k-sparse codebook is then defined as follows.

Definition 4.5. *A codebook $\mathcal{F} = \{\mathbf{f}_i\}_{i=1}^N$ for $\mathbb{C}^m$ is called k-sparse if the sum of non-zero elements in the real and imaginary parts of the frame synthesis matrix $\mathbf{F}$ is k. Here, $0 \leq k \leq 2mN$.*

Note that the definition of k-sparsity is with respect to the standard basis. To quantify the sparsity, the density (with respect to sparsity) of a frame is defined as follows.

Definition 4.6. *The density κ of a frame is the ratio of non-zero components and the total number of components in the frame. In particular, the density is given by $\kappa = \frac{k}{2mN}$.*

Table 4.3: Complexity Savings Using Proposed Method

N_t	$N = 8$	$N = 16$	$N = 64$	$N = 128$
2	50%	50%	50%	50%
4	39%	26%	17%	16%
8	23%	15%	8%	7.7%

New sparsity properties are now proven for codebooks with elements that are the N-th roots of unity. An example is the Fourier codebook [Love et al., 2003]. In particular, the following proposition is developed.

Proposition 4.2. *For quantization codebooks $\mathcal{F}$ with elements that are the N-th roots of unity for a BS employing $N_t = 2$ antennas, there exists a unitary transformation $\mathbf{U}$ and real numbers $\theta_i \in [0, 2\pi)$, $i = 1, \ldots, N$ such that the synthesis matrix $\mathbf{U}[e^{j\theta_1}\mathbf{f}_1, \ldots, e^{j\theta_N}\mathbf{f}_N]$ has elements that are either purely real or purely imaginary, and $\mathbf{U}$ is constructed using the Householder construction in Table 4.1. Moreover, the density is $\kappa = \frac{2N-1}{4N}$.*

Proof: See Appendix B.2 for a constructive proof. ■

Proposition 4.2 shows that a reduction in search complexity of over 50% is possible for codebooks with elements that are the N-th roots of unity. This is achieved by not requiring the computation of multiplications and additions for both real and imaginary parts in the codeword elements.

Sparsity properties of codebooks for $N_t > 2$ depend on the particular vectors in the codebook. Extensive codebook constructions of the type in Proposition 4.2 have revealed that complexity savings can also be obtained for $N_t > 2$. In particular, Table 4.3 shows the complexity savings using the proposed method for different N_t and N. Observe that for $N_t = 2$, the complexity savings are at least 50% compare to the standard Fourier base codebook [Love et al., 2003], which confirms the result in Proposition 4.2. For $N_t > 2$, observe that significant complexity savings are still obtained; however, these complexity savings depend on the size of the codebook.

4.4 Rate-Outage Lower Bound and Base Codebook Design

Previous analysis on existing feedback schemes for multiuser MIMO has been under the assumption of random codebook generation. Random codebooks do not suffer from problems of different users selecting the same codeword [Ding et al., 2007]. Unfortunately, they are impractical. The proposed practical constructions also do not suffer the "same codeword problem".

In contrast, the existing practical common codebook scheme leads to a loss in the sum-rate and that is why analysis for the common codebook scheme is provided in this section. To understand the benefit of the two constructions, analysis of this rate loss in terms of the common codebook scheme is required. To do this, the outage probability lower bound for the common codebook scheme is calculated, which is also the basis of an approximate lower bound for the sum-rate of the scheme in the high SNR regime. In particular,

$$E[R_{CC}] \approx (1 - P_o)E[R], \tag{4.19}$$

where R_{CC} is the sum-rate of the common codebook scheme, P_o is the outage probability and R is the sum-rate of the scheme. Hence, $E[R] \gtrapprox E[R_{CC}]$. Moreover, by minimizing P_o, $E[R_{CC}]$ is also maximized and consequently the approximate lower bound of the sum-rate of the scheme.

The key result in this section is the optimization of the base codebook (i.e. the codebook used to generate all codewords for each user). The base codebook derivation is based on the rate-outage lower bound analysis. It is this optimized base codebook that is used for simulations in this section for both the common codebook and the GR construction.

It is assumed that the network supports real-time services in which rank adaptation is not acceptable, even when several users send back identical codewords that results in rank deficiency. This allows for minimum rate requirements for each user to be met. As a consequence, rate-outage occurs if the quantized channel matrix is non-invertible. Denote the sum-rate outage probability as $P_o = \Pr(R \leq R_0)$, where R_0 is a fixed rate-

threshold. By conditioning on the invertibility of the quantized channel matrix,

$$\begin{aligned} P_o &= \Pr(\hat{\mathbf{H}} \text{ invertible}) \Pr(R \leq R_0 | \hat{\mathbf{H}} \text{ invertible}) + \Pr(\hat{\mathbf{H}} \text{ not invertible}) \\ &\geq \Pr(\hat{\mathbf{H}} \text{ not invertible}). \end{aligned} \tag{4.20}$$

Observe that when $R_0 \approx 0$, the bound is tight at high SNR as $\Pr(R \leq R_0 | \hat{\mathbf{H}} \text{ invertible}) \approx 0$ in this regime.

To determine the minimum value of $\Pr(\hat{\mathbf{H}} \text{ not invertible})$, assume that each possible quantized channel matrix has an N_t-dimensional column space. This is to avoid any unnecessary rank deficiency due to linearly dependent columns. Also, note that the codebook minimizing $\Pr(\hat{\mathbf{H}} \text{ not invertible})$ is equivalent to the codebook maximizing $\Pr(\hat{\mathbf{H}} \text{ invertible})$.

The following constrained non-convex optimization problem is constructed to obtain the probability of selecting each codeword to minimize the probability that the quantized channel matrix is not invertible for isotropic channel vectors. Note that this is equivalent to obtaining the probability of selecting the codewords to maximize the probability that the quantized channel matrix is invertible, as these events are complements of each other. The optimization problem is constructed by determining the sets of distinct codewords, which are denoted by $Z_j \subset \{1, \ldots, N\}$ with $|Z_j| = N_t$, $j = 1, \ldots, \sigma$. Observe that the probability that a given set of distinct codewords is selected is given by $\prod_{i \in Z_j} p_i$, where p_j is the probability that the j-th codeword is selected. The total probability that the quantized channel matrix is invertible is then given by $\sum_{j=1}^{\sigma} \prod_{i \in Z_j} p_i$, where $\sigma = \binom{N}{N_t}$ is the total number of combinations of distinct codewords. The required optimization problem is then

$$\begin{aligned} \max_{p_i,\, i=1,\ldots,N} \quad & \sum_{j=1}^{\sigma} \prod_{i \in Z_j} p_i \\ \text{subject to} \quad & 0 \leq p_i, \; \forall \, i = 1, 2, \ldots, N \\ & \sum_{i=1}^{N} p_i = 1. \end{aligned} \tag{4.21}$$

In this case, following proposition holds.

Proposition 4.3. *For the non-convex optimization problem in (4.21), the KKT conditions are*

necessary and sufficient for global optimality and the optimal probabilities, p_i^*, $i = 1, \ldots, N$, are given by

$$p_1^* = \cdots = p_N^* = \frac{1}{N}. \tag{4.22}$$

Proof: See Appendix B.3. ■

Using a simple counting argument, the bound in (4.20) and Proposition 4.3, the rate-outage probability for MU-MIMO using ZF precoding is then lower bounded by

$$\begin{aligned} P_o &\geq 1 - \Pr(\hat{\mathbf{H}} \text{ invertible}) \\ &\geq 1 - \frac{N!}{(N-N_t)!N^{N_t}}. \end{aligned} \tag{4.23}$$

The second inequality in (4.23) holds as it is obtained by using the optimal probability that each codeword is selected (obtained by solving the optimization problem in (4.21)) to compute the probability that the quantized channel matrix $\hat{\mathbf{H}}$ is invertible.

The first key analytical result in this section provides the benefit of using the constructions instead of the common codebook scheme. In particular, the average sum-rate loss due when employing a common codebook compared with employing different codebooks is approximately lower bounded by

$$R_L \gtrapprox \left(1 - \frac{N!}{(N-N_t)!N^{N_t}}\right) E[R], \tag{4.24}$$

where $E[R]$ is the expected rate. The bound follows from (4.23). As such, the relative sum-rate loss if a system used a common codebook compared with the proposed codebooks that do not suffer this loss is approximately

$$\left(1 - \frac{N!}{(N-N_t)!N^{N_t}}\right). \tag{4.25}$$

The second key analytical result is the base codebook that can be used as the basis for the new proposed construction as indicated in Step 1) of Table 4.2. This base codebook minimizes the outage probability lower bound for isotropic channels. Before proving the result, an ETF is defined as follows.

Definition 4.7. *A finite frame for* $\mathbb{C}^{N_t}$ *[Strohmer and Heath, 2003],* $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N]$ *with* $\|\mathbf{f}_i\|^2 = 1$, $i = 1, \ldots, N$ *is an ETF if the following conditions are satisfied:*

1. $\mathbf{F}\mathbf{F}^\dagger = \frac{N}{N_t}\mathbf{I}$,

2. $|\mathbf{f}_i^\dagger \mathbf{f}_j|^2 = \frac{N - N_t}{N_t(N-1)}$, $i \neq j$.

An important property of ETFs is that they can only exist for $N \leq N_t^2$ [Strohmer and Heath, 2003]. This means that base codebooks generated using this method are limited to at most N_t^2 codewords.

To show that certain types of ETFs have equally likely codewords and hence minimize the rate-outage lower bound in (4.23), a result in [Mondal et al., 2007] is generalized for quantizers on the real projective space $\mathbb{R}\mathrm{P}^{N_t-1}$ to the complex projective space $\mathbb{C}\mathrm{P}^{N_t-1}$. This is given in Theorem 4.2.

Theorem 4.2. *Let* $\mathcal{F} = [\mathbf{f}_1, \mathbf{U}_2\mathbf{f}_1, \ldots, \mathbf{U}_N\mathbf{f}_1]$ *be an ETF with* N *elements and dimension* N_t*, and each element of* $\boldsymbol{H}$ *is distributed as* $h_{ij} \sim \mathcal{CN}(0,1)$*. Further, let* $\mathbf{F}$ *be a frame for* $\mathbb{R}^d$*, or* $\mathbf{U}_i\mathbf{U}_j = e^{j\zeta_{ij}}\mathbf{U}_{ij}$ $1 \leq i, j \leq N$, $\zeta_{ij} \in [0, 2\pi)$*, where* $\{\mathbf{U}_i\}_{i=1}^N$ *forms a group. If all potential quantized channel matrices* $\hat{\mathbf{H}}$ *have an* N_t*-dimensional column space, then each quantization codeword in* $\mathcal{F}$ *is equally likely and* $\mathcal{F}$ *minimizes the rate-outage lower bound in (4.23).*

Proof: See Appendix B.4. ■

For the ETFs of the form specified in Theorem 4.2 it is necessary to find the $\mathbf{U}_i$ and $\mathbf{f}_1$ that satisfy the constraints in Theorem 4.2. Previous work in the theory of ETFs has provided methods to find these for a range of N. An important case is when $N = N_t^2$, which has been extensively studied in the quantum information theory literature. In particular, explicit cases of $\mathbf{U}_i$ and $\mathbf{f}_1$ are available in [Scott and Grassl, 2010] for dimensions up to $N_t = 67$.

Since $\mathcal{F}$, generated using Theorem 4.2, minimizes the rate-outage lower bound it is an excellent choice for the base codebook when $N \leq N_t^2$. This construction is used in the simulations in the next section for the cases $N \leq N_t^2$.

4.5 Simulation Results

In this section, Monte Carlo simulations are performed to verify the performance of the proposed codebook constructions. The simulations show that the proposed codebooks perform comparably with the random codebook proposed in [Jindal, 2006] with reduced complexity. The simulation results are shown to be consistent with analysis. In all simulations the total power is $P = 1$, and $\rho = 1/\sigma^2$.

In Fig. 4.2, the sum rate of an $N_t = K = 2$ is compared with ρ for a system with each user using a different quantization codebook and a base codebook obtained using the Lloyd algorithm [Xia et al., 2005]. Three constructions are used to obtain a different codebook at each user: 1) the GR construction; 2) the random unitary transform; and 3) the common codebook at each user. In the GR construction, the set of unitary matrices in Example 2 from the group $G_{6,-1}$ is used. Observe that the GR construction reduces the storage requirements compared with the random construction, while achieving the same performance for all codebook sizes.

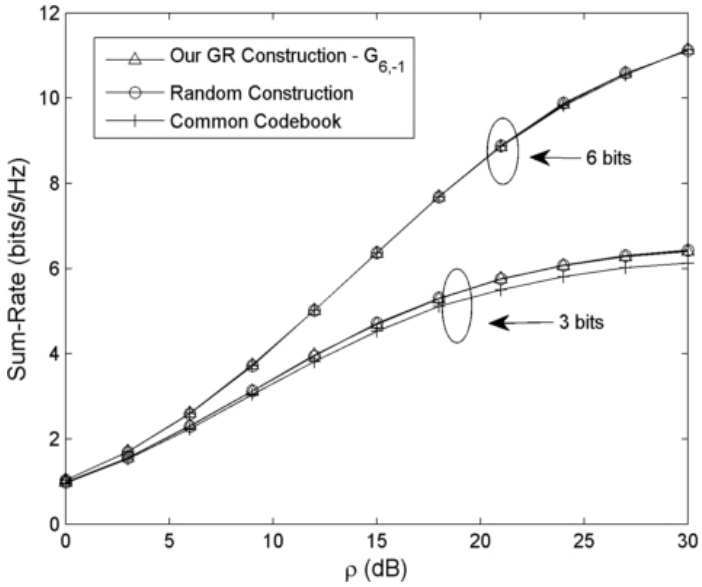


Figure 4.2: Sum-rate with $N_t = K = 2$ using the Lloyd algorithm base codebook and different transformation constructions: the GR construction based on the $G_{6,-1}$ in Example 2, the random unitary transform [Ding et al., 2007] and a common codebook at each user.

In Fig. 4.3, the sum-rate of an $N_t = K = 4$ is compared with ρ for a system where each user using a different quantization codebook and the base codebook generated using the

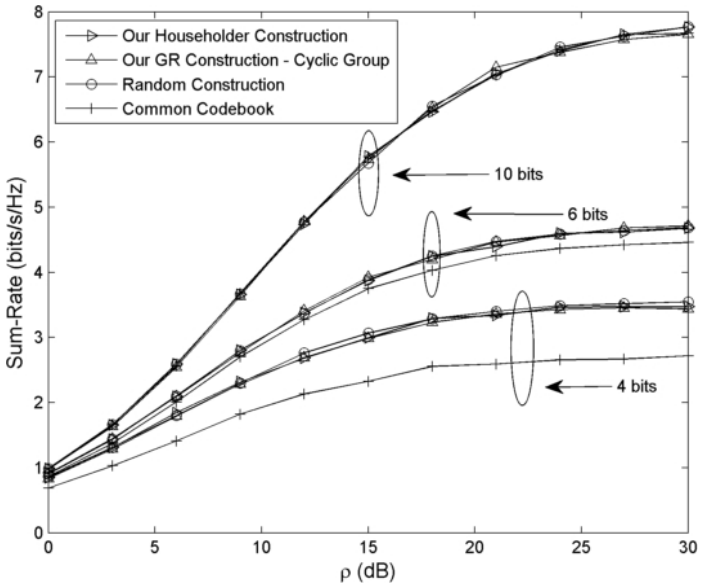


Figure 4.3: Sum-rate with $N_t = K = 4$ using the Lloyd algorithm base codebook and different transformation constructions: the GR construction based on a cyclic group, the Householder construction, the random unitary transform [Ding et al., 2007] and a common codebook at each user.

Lloyd algorithm [Xia et al., 2005]. Four constructions are used to obtain the codebooks at each user: 1) the Householder construction, 2) the GR construction, 3) the random unitary transform and 4) a common codebook at each user. In the GR construction the cyclic group with $n = 7$ and $k_1 = 1, k_2 = 2,\ k_3 = 3,\ k_4 = 5$ is used (see Example 1 for the detailed construction). Observe that the Householder and GR constructions reduce the storage requirements compared with the random construction, while achieving the same performance for all codebook sizes. Note the difference from Fig. 4.2 in that 10 bits are required for the common codebook to perform comparably with the random construction. This is a huge codebook, which means that the Householder or GR constructions are necessary for practical codebook sizes. In the $N_t = K = 2$ case, only 3 bits were required. This shows that as N_t increases, so does the number of codewords such that the common codebook can be used with negligible degradation compared with the transformed codebook constructions.

In Fig. 4.4, the sum-rate of network with $N_t = 4$ and $K = 2$ is plotted. Two base codebooks are used: the optimal base codebook (4 bits); and the Lloyd algorithm base

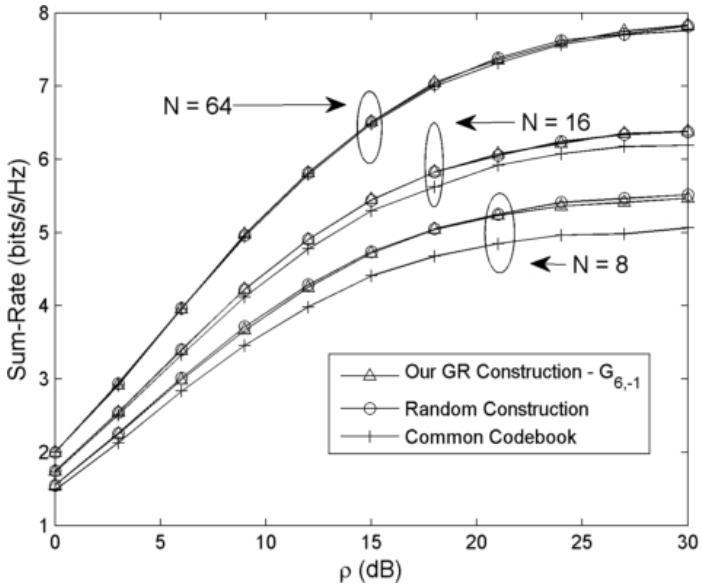


Figure 4.4: Sum-rate with $N_t = 4$, $K = 2$ using the Lloyd algorithm base codebook (3 and 6 bits), ETF base codebook (4 bits) and different transformation constructions: the GR construction based on a cyclic group, the random unitary transform [Ding et al., 2007] and a common codebook at each user.

codebook (6 and 10 bits). Three constructions are used to obtain the codebooks: 1) the GR construction, 2) the random unitary transform and 3) a common codebook. As for the case where $N_t = K = 4$, the GR construction performs comparably with the random construction in [Ding et al., 2007], while storing only one diagonal unitary matrix instead of at least 200 for the random construction in LTE cells.

Fig. 4.5 shows the outage probability where each user has the same codebook and uses the Lloyd algorithm [Xia et al., 2005] to generate the base codebook. The figure shows that the bound holds for all simulated codebook sizes and numbers of transmit antennas. Observe that the lower bound (4.23) is tight at high ρ for a rate-threshold of 0.01 bits/channel use.

Fig. 4.6 shows the probability that a target rate is met for each user. The base station has $N_t = 2$ antennas, which can transmit to one or two users depending on whether minimum rate constraints are met. When a per-user rate constraint for either user is not met using two user transmission, then the user with the highest rate using single-user beamforming is serviced. In the figure, the target rate for user 1 is $R_{0,1} > 0.1$ and the target

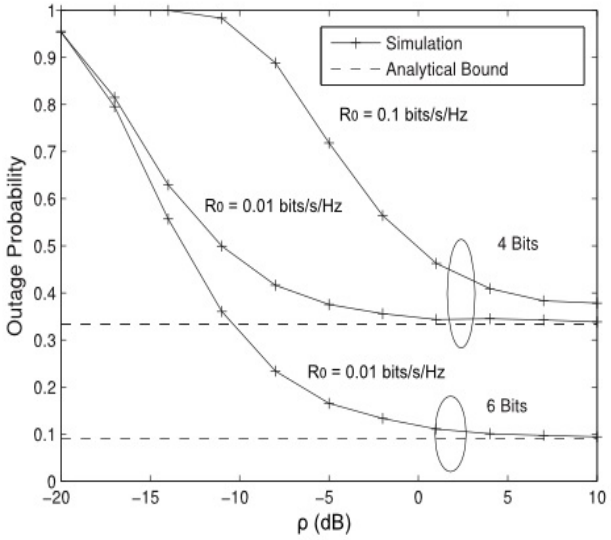


Figure 4.5: Outage probability bound from (4.23) with $N_t = K = 4$ and rate-threshold R_0 bits/s/Hz using the Lloyd algorithm with a common codebook at each user.

rate for user 2 is $R_{0,2} > 0.2$. The per-user outage probability is $\Pr(R_i < R_{0,i})$. The base codebook is constructed using the Lloyd algorithm with a size of 2 bits. Observe that using the Householder construction can achieve up to a 15% improvement in the probability that the rate target is met for ρ above 15 dB. This shows that even if scheduling is employed to avoid the rank deficiency, there is a reduction in the probability that a given user can meet its data rate target compared with the proposed Householder construction.

4.6 Conclusion

Two new codebook constructions were proposed to obtain different codebooks at each user. The Householder construction employed the Householder transform and constructed unique codebooks at each user with no storage requirements, in addition to the base codebook. The second construction was based on unitary representations of finite groups, which is applicable to any base codebook. The search complexity of the Fourier base codebook was shown to be reduced using the proposed Householder construction by up to 50% compared with the standard Fourier base codebook. To compare the constructions with the standard case of a common codebook at each user, a lower bound on

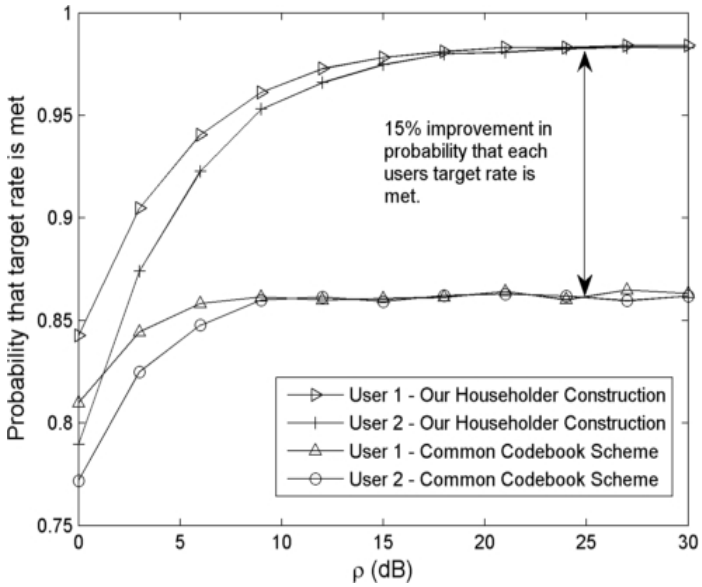


Figure 4.6: Probability that user target rates $R_1 > 0.1$ and $R_2 > 0.2$ are met with $N_t = 2$ and at most $K = 2$ users are serviced using the Lloyd algorithm base codebook, with the Householder construction and the common codebook scheme.

the rate-outage probability was derived. This lead to a new optimal base codebook construction, which can be used with the constructions to obtain different codebooks at each user. Simulations were performed that showed both the proposed constructions significantly reduce the storage requirements compared with the random construction, while achieving the same sum-rate performance for a wide range of popular base codebooks.

Chapter 5

Structured Limited Feedback Codebook Design for Coordinated Multipoint

An important new feature present in CoMP is that variable numbers of BSs can service a given user. This poses a new problem for beamforming that is not present in single-cell operation: the quantization codebook must support a variable dimension codebooks, with the dimension corresponding to the number of BSs employed. This is a problem that has not appeared in previous releases of LTE. In this chapter, a low complexity structured codebook is proposed that has linear complexity in both the codebook size and dimension. As such, the variable dimension codebooks are readily accommodated and the codebook can be constructed online as the number of transmitting BSs varies. A new method is also proposed to store optimal structured codebooks–in the sense of the Grassmannian criterion–of variable dimension by exploiting properties of cyclic difference sets.

5.1 Introduction

AS for the single-cell MU-MIMO networks in Chapter 4, CSI is required for beamforming in CoMP using FDD. This is achieved by using a low-rate feedback link from each user to the BSs [Jindal, 2006]. An effective approach to CSI feedback in FDD networks is to employ codebook-based quantization [Egan et al., 2011, Jindal, 2006, Heath et al., 2009, Love et al., 2003, Chae et al., 2008, Ryan et al., 2009], which is currently standardized for single-cell networks. In codebook-based quantization, a set of quantized channel vectors are stored at each user. The user then chooses the codeword that best quantizes the actual channel vector.

There are two key differences between quantization for single-cell networks and CoMP.

The first is that CSI feedback is potentially required for a large number of links due to a large number of cooperating BSs. The second is that a variable number of BSs can transmit to each user, requiring variable codebook dimension.

Current approaches to the problem of quantization codebook design for CoMP have focused on random constructions. Random approaches are undesirable as both storage and search requirements are prohibitive in contrast with structured approaches such as the Fourier codebook [Love et al., 2003]. In [Cheng et al., 2010], a general codebook construction criterion based on the chordal distance was proposed for network MIMO exploiting per-cell codebooks. This technique decomposed the problem of globally determining the optimal codebook for all BSs to a joint codebook composed of codewords obtained by only considering individual cells. Unfortunately, specific constructions were not considered and analysis was performed only for random constructions. In [Hou and Yang, 2011] and [Su et al., 2011], bit allocations for different channels and codeword selection algorithms were proposed, respectively, when per-cell codebooks are employed. As in [Cheng et al., 2010], the problem of constructing structured codebooks was not considered.

In this chapter, a low-complexity structured codebook construction is proposed for a CoMP network. In particular, the problem of prohibitive complexity of variable dimension Fourier codebook construction is overcome, while preserving the high-performance of the Fourier (or Grassmannian) codebook [Love et al., 2003]. Our algorithm to construct variable dimension codebooks has a complexity linear in both codebook size and dimension (which corresponds to the total number of antennas used by the transmitting BSs). This is a significant reduction compared with the exhaustive search technique, which has exponential complexity. Our construction is based on a new recursive bound on the maximum distance between codewords. This alleviates the problem of computing exponential sums, which are the main difficulty in analyzing and designing optimal Fourier codebooks.

Next, a new storage method is proposed for Fourier codebooks, which are also equiangular tight frames (ETFs)–an optimal class of codebooks. It is shown that ETFs of different dimensions can be constructed using a single set of coefficients and exploit this, as

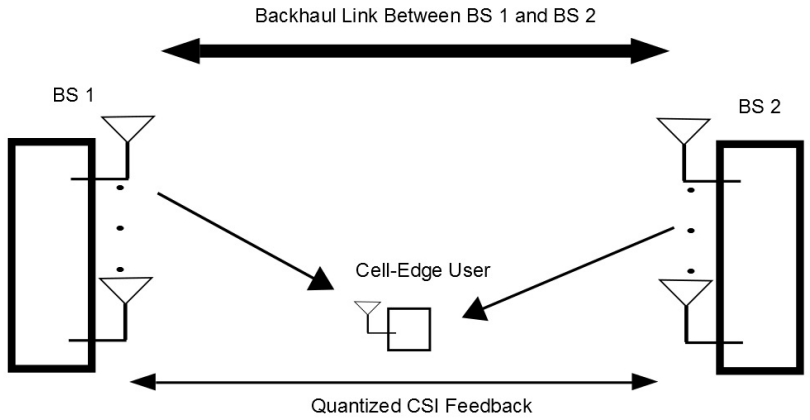


Figure 5.1: CoMP network with feedback.

yet unused, property to develop the proposed storage method. A key feature of the approach is that only a single set of coefficients are required to construct Fourier codebooks when varying numbers of BSs can transmit. This means that the storage requirements are comparable to codebooks used in LTE, while providing for optimality.

Simulations are used to show that the proposed low-complexity construction performs comparably with the exhaustive search, with linear complexity in the size of the codebook and dimension of the codewords. This means that codebooks of varying size and dimension can be constructed online as there are changes over time of the number of coordinated BSs or signal-to-interference and noise ratios (SINR) of the links.

5.2 System Model

Consider a CoMP network (see Fig. 5.1, with K BSs and a single-antenna user. BS i has a total of N_i transmit antennas, and is at a distance R_i from the user. Assume that the data for the user and the CSI from the BSs to the user is sent to a central unit (CU) via error-free and zero-latency backhaul links. The channel is block fading and the network operates using FDD, which means that reciprocity in the channel cannot be exploited in

sharing CSI between the user and each BS.

In each time slot, the user that is serviced by the coordinated BSs perfectly estimates the channel between itself and each BS. The user then forms a channel vector

$$\mathbf{h}^\dagger = [\mathbf{h}_1, \ldots, \mathbf{h}_d], \tag{5.1}$$

where $\mathbf{h}_i^\dagger$ is the channel between the user and BS i, and $d = \sum_i N_i$ is the total number of antennas employed by all the coordinated BSs in the time slot. The user then computes the channel shape

$$\tilde{\mathbf{h}} = \mathbf{h}/\|\mathbf{h}\|. \tag{5.2}$$

At the user and the CU, a codebook consisting of d-dimensional quantization codewords $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N]$, $\|\mathbf{f}_i\|^2 = 1$, $i = 1, 2, \ldots, N$, of size N is stored. Once the user has estimated the channel vector, it computes the inner product between each codeword and the channel shape to determine which codeword index to feedback to the CU. In particular,

$$i^* = \arg \max_{i=1,\ldots,N} |\tilde{\mathbf{h}}^\dagger \mathbf{f}_i|^2, \tag{5.3}$$

is the codeword index fed back to the CU. The coefficients of the codeword i^* are then shared with each BS to perform beamforming.

The throughput using distributed beamforming is given by

$$R = \log_2 \left(1 + \frac{|\mathbf{h}^\dagger \mathbf{P} \mathbf{f}_{i^*}|^2}{N_0}\right), \tag{5.4}$$

where $P = \text{diag}\{[P_1, \ldots, P_d]\}$ is the transmit power matrix and N_0 is the noise power. Note that (5.4) has a similar form to single-cell MISO transmission. The difference lies in the fact that the average channel gains in $\mathbf{h}$ may be different for each element h_i due to the fact that the distance between the user and each BS may vary. The transmit signal-to-noise ratio (SNR) for BS i is given by P_i/N_0.

5.3 Codebook Design

In practical wireless systems, random quantization codebooks are infeasible due to high storage and search complexity. In order to alleviate these problems, structured codebooks are required. Structured codebook construction has been extensively studied for single-cell networks (see [Egan et al., 2011, Ryan et al., 2009, Heath et al., 2009, Love et al., 2003, Jindal, 2006] and the references therein). In this section, the codebook design criteria is shown to maximize the throughput for single-cells can in fact be directly translated to CoMP networks and as such the Fourier codebook is a desirable choice.

First observe that CoMP networks employ a CU to compute the beamforming vector in a centralized fashion. As such, CoMP networks can be regarded as a generalization of standard single-cell beamforming. Moreover, since only the channel shape is required in order to perform beamforming, there is no difference in the computation of the beamforming vector for CoMP networks and single-cell networks that only employ the channel shape. As a result, the design criteria for quantization codebooks employed in single-cell networks can be exploited.

A common design criterion for quantization codebooks in the MIMO broadcast channel is the Grassmannian criterion [Love et al., 2003], which minimizes the chordal distance between codewords. This is given by

$$\min \max_{i \neq j} |\mathbf{f}_i^\dagger \mathbf{f}_j|^2. \tag{5.5}$$

Unfortunately, the optimal codebook using the Grassmannian criterion is not known or unstructured for most codebook sizes.

In [Egan et al., 2011], a coarser criterion known as the expected square correlation (ESC) was proposed for the MIMO broadcast channel with zero-forcing (ZF) precoding. As the beamforming used to obtain the throughput in (5.4) is a special case of ZF, the ESC can be used to find classes of codebooks with high throughput. The Grassmannian criterion can then be applied to optimize over the class of codebooks, which is significantly easier than the original problem of optimizing over all possible codebooks. The

ESC criterion is given by

$$\max_{\mathbf{F}} E\left[1 - |\mathbf{f}_i^\dagger \mathbf{f}_j|^2\right]. \tag{5.6}$$

A key property of the ESC criterion is that it is minimized by codebooks with group structure. That is, the codebooks constructed as

$$\mathbf{F} = [\mathbf{U}_1\phi, \ldots, \mathbf{U}_N\phi], \tag{5.7}$$

where $\{\mathbf{U}_i\}_{i=1}^N$ is an algebraic group of unitary matrices and $\phi \in \mathbb{C}^d$ with $|\phi|^2 = 1$. An important class of ESC-minimizing codebooks are known as Fourier (or Grassmannian) codebooks, where the group of unitary matrices forms a cyclic group. Recall that the Fourier codebook also played an important role in the construction in Chapter 4. As noted in Section 4.2, the Fourier codebook can be easily constructed using the direct sum to obtain a reducible representation of a cyclic group generated by

$$\begin{aligned}\mathbf{U} &= \chi_1 \oplus \cdots \oplus \chi_d \\ &= \begin{pmatrix} e^{2\pi j k_1/N} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & e^{2\pi k_d/N} \end{pmatrix},\end{aligned} \tag{5.8}$$

where $\oplus$ is the direct sum and $\chi_i = e^{2\pi j k_i/N}$ for some $1 \leq k_i \leq N-1$. Hence, each codeword in the Fourier codebook is given by $\mathbf{f}_l = \mathbf{U}^l \mathbf{Œ}$, $l = 0, \ldots, N-1$, where $\mathbf{Œ} = 1/\sqrt{d}[1, 1, \ldots, 1]^T$.

Unfortunately, there is no known exact low complexity search algorithm to obtain the coefficients $k_1, \ldots, k_N$. Instead, an exhaustive search of exponential complexity or a random search is usually employed [Hochwald et al., 2000]. To reduce the complexity, an approximate low complexity algorithm is proposed in the next section to obtain the coefficients.

5.4 Proposed Low Complexity Construction

In this section, a low complexity construction of the Fourier codebook is proposed based on the Grassmannian criterion. A key problem in CoMP that does not appear in the single cell case is that codebooks must be constructed for variable numbers of BSs and codebook sizes, which changes the dimension of the codebook. Our codebook construction has complexity of order $O(Nd)$ in contrast to the standard exhaustive search approach, which has exponential complexity.

Our proposed construction is based on an iterative procedure, which generates one coefficient at a time. In order to develop the construction, a new criterion based on the Grassmannian criterion is required in order to generate the coefficient at each step.

The first step is to observe that the norm-squared inner product–the basis of the Grassmannian criterion–between two codewords for the $d = 2$ codebook is

$$\begin{aligned} |\mathbf{f}_{l_q}^{\dagger}\mathbf{f}_{l_r}|^2 &= |\frac{1}{2}\sum_{m=1}^{2} e^{-2\pi j k_m (l_l - l_k)/N}|^2 \\ &= \frac{1}{4}\left(2 + e^{2\pi j (l_l - l_k)(k_1 - k_2)/N} + e^{2\pi j (l_r - l_q)(k_2 - k_1)/N}\right) \\ &= \frac{1}{2} + \frac{1}{4}(\psi_1^* \psi_2 + \psi_1 \psi_2^*), \end{aligned} \tag{5.9}$$

where $\psi_i = e^{2\pi j (l_l - l_k) k_i / N}$. Denote l_1, l_2 as the pair of codewords that maximizes (5.9), with $l_1 \neq l_2$.

Now, the norm-squared inner product for the case where $d = 3$ and codebook size M is given by

$$|\mathbf{f}_{l_q}^{\dagger}\mathbf{f}_{l_r}|^2 = |\frac{1}{3}\sum_{m=1}^{3} e^{-2\pi j k_m (l_r - l_q)/M}|^2. \tag{5.10}$$

The next step is to choose $M(l_1 - l_2) = N(l_l - l_k)$ for some $l_l, l_k \in \{1, 2, \ldots, M\}$ (recall that l_1, l_2 are the pair of codewords that maximize (5.9)). This allows (5.10) to be written in terms of ψ_1, ψ_2 and hence the coefficients k_1, k_2 to be obtained from (5.9). In particular,

$$\max_{l_q \neq l_r} |\mathbf{f}_{l_q}^{\dagger}\mathbf{f}_{l_r}|^2 = |\frac{1}{3}\sum_{m=1}^{3} e^{-2\pi j k_m (l_1 - l_2)/N}|^2$$

$$\geq \frac{1}{9}\Big(3+\psi_1^*\psi_2+\psi_1\psi_2^*+e^{2\pi j(k_1-k_3)(l_1-l_2)/N} + e^{2\pi j(k_2-k_3)(l_1-l_2)/N}+e^{2\pi j(k_3-k_1)(l_1-l_2)/N}+e^{2\pi j(k_3-k_2)(l_1-l_2)/N}\Big), \quad (5.11)$$

where $\psi_1^*\psi_2+\psi_1\psi_2^*$ is the same as in (5.9). The bound follows by noting that the inner product of codewords l_1 and l_2 must be less than or equal than the maximum pair.

We now consider $d>3$, which means that the dimension of the codewords is $d>3$. More generally, the following bound holds as the dimension is increased further

$$\max_{l_q\neq l_r}|\mathbf{f}_{l_q}^\dagger\mathbf{f}_{l_r}|^2 \geq \frac{1}{d^2}\left(d+\sum_{i=1}^{d}\sum_{j=1,\ j\neq i}^{d}\psi_j\psi_i^*\right), \quad (5.12)$$

where $\psi_1=1$. As the dimension of the codebook is increased, the new characters can be computed as

$$\psi_i = e^{2\pi j k_i(l_1-l_2)/N}. \quad (5.13)$$

Observe that the contribution in (5.12) of each new term corresponding to a unit increase in the codebook dimension is given by

$$\sum_{i=1}^{d+1}\psi_i\psi_{d+1}^*+\psi_i^*\psi_{d+1}. \quad (5.14)$$

This sum provides a method for choosing k_{d+1}. In particular, k_{d+1} is obtained by solving the optimization problem

$$\min_{k_{d+1}=1,2,\ldots,N}\ \sum_{i=1}^{d+1}\psi_i\psi_{d+1}^*+\psi_i^*\psi_{d+1} \quad (5.15)$$

The proposed construction for the Fourier codebook is constructed as follows.

1. Obtain the coefficients k_1,k_2 using the Grassmannain criterion and exhaustive search for the $d=2$ codebook with N codewords. The exhaustive search is feasible due to the low dimension of the codewords.

2. Set $d=d+1$ and choose l_k,l_l in (5.13) such that $M(l_1-l_2)=N(l_k-l_l)$, where M

is the size of the new codebook of dimension d.

3. Compute the characters ψ_i in (5.13) and solve (5.15) based on the bound in (5.12) to obtain k_d.

4. Repeat from Step 2) until the desired codebook dimension is reached.

In contrast with the standard exhaustive search, this approach only requires d searches over N integers instead of a single search over N^d integers. As such, the complexity of obtaining the coefficients using (5.15) to construct the codebook is of order $O(Nd)$. The performance of the two approaches is compared using simulations, where the proposed low complexity construction is shown to perform comparably with the exhaustive search approach.

Note that the low complexity codebook construction can be employed as an initial codebook in other codebook constructions. In particular, it is directly applicable in the schemes proposed in [Heath et al., 2009, Egan et al., 2013a].

5.5 ETF-Based Storage Method

In this section, a method is proposed to construct a sequence of codebooks with different dimensions using only one set of coefficients, while preserving the optimality of each codebook. This is achieved by using Fourier codebooks that are also equiangular tight frames–codebooks that optimize the Grassmannian criterion. More precisely, an ETF is a codebook that satisfies $\max_{i\neq j} |\mathbf{f}_i^\dagger \mathbf{f}_j|^2 = \frac{N-d}{d(N-1)}$. In fact, this is the minimum achievable value of $\max |\mathbf{f}_i^\dagger \mathbf{f}_j|^2$ as it achieves the Welch bound [Xia et al., 2005].

Before describing the proposed method, a result from frame theory and combinatorics is required. In particular, the notion of a cylic difference set is required, which is defined as follows.

Definition 5.1. *A subset* $\mathbf{u} = \{u_1, \ldots, u_d\}$ *of* $\mathbb{Z}_N$ *is called a* (N, d, λ) *difference set if the* $d(d-1)$ *differences*

$$(u_k - u_l) \mod N,\ k \neq l \tag{5.16}$$

take all possible values $1, 2, \ldots, N-1$*, with each value exactly* λ *times.*

The precise link between Fourier codebook, ETFs and difference sets was proven in [Xia et al., 2005] and is summarized in the following theorem.

Theorem 5.1 ([Xia et al., 2005])**.** *The Fourier codebook with coefficients* $\mathbf{k} = [k_1, \ldots, k_d]$ *is an ETF if and only if* $\mathbf{k}$ *is an* (N, d, λ) *cyclic difference set.*

The reduced-storage method is now outlined for constructing variable dimension codebooks using only a single set of coefficients. Our method uses an important, but so far neglected, property of cyclic difference sets. In particular, there is a class of cyclic difference sets that contain subsets that are also cyclic difference sets, which is referred to as the *subset property*. A partial classification of cyclic difference sets with the subset property is developed in [Jungnickel and Tonchev, 1999]. Two examples are provided to illustrate this idea:

Example 1:

1. $d = 3,\ N = 7$: $\mathbf{k} = \{1, 2, 4\}$;
2. $d = 7,\ N = 15$: $\mathbf{k} = \{1, 2, 4, 0, 5, 8, 10\}$.

Example 2:

1. $d = 4,\ N = 13$: $\mathbf{k} = \{0, 1, 3, 9\}$;
2. $d = 13,\ N = 40$:
   $\mathbf{k} = \{0, 1, 3, 9, 5, 15, 22, 25, 26, 27, 34, 35, 38\}$.

Our proposed method proceeds as follows:

1. Choose the dimensions and the number of codewords that are required for each codebooks.
2. Construct the cyclic difference sets for each codebook–each cyclic difference set is equivalent to the set of coefficients $k_1, \ldots, k_d$ for that dimension. If the subset property holds, order the largest cyclic difference set such that the first elements correspond to smaller cyclic difference sets. This is illustrated in both examples above.

3. When a total d transmit antennas are employed, choose the first d elements of the largest cyclic difference set.

4. Construct the codebook using (5.8).

In order to employ the method for the largest possible class of cyclic difference sets, the performance of the codebook should not be changed by the order of the cyclic difference sets within the largest cyclic difference set. This is proved in the following proposition.

Proposition 5.1. *Let* $\mathbf{k}$ *be the coefficients for a Fourier codebook* $\mathbf{F}$*, and* $\mathbf{k}'$ *be a permutation of* $\mathbf{k}$ *that generates the codebook* $\mathbf{F}'$*. Then,*

$$\max_{i \neq j} |\mathbf{f}_i^\dagger \mathbf{f}_j|^2 = \max_{k \neq l} |(\mathbf{f}'_k)^\dagger \mathbf{f}'_l|^2, \tag{5.17}$$

where $\mathbf{f}_i$*,* $\mathbf{f}_j$ *are in* $\mathbf{F}$ *and* $\mathbf{f}'_k$*,* $\mathbf{f}'_l$ *are in* $\mathbf{F}'$*.*

Proof: Observe that the permutation operator is unitary. As such, the codewords in $\mathbf{F}'$ can be written as

$$\mathbf{f}'_i = \mathbf{U}_\pi \mathbf{f}_j, \tag{5.18}$$

where $\mathbf{f}_j$ is in $\mathbf{F}$ and $\mathbf{U}_\pi$ represents the permutation. Since the unitary transformation is an isometry, the proposition follows. ■

A useful property of difference sets is that the complement is also a difference set. This means that new ETFs can be easily constructed from pre-existing ones. In [Xia et al., 2005], the following property was proven.

Proposition 5.2. *Let* $\mathbf{k}$ *be the coefficients of an ETF that form a cyclic difference set over* $\mathbb{Z}_N$*. Then,* $\mathbf{k}' = \mathbb{Z}_N \setminus \mathbf{k}$ *is also a cyclic difference set corresponding to an ETF.*

As a consequence, the following useful proposition is proven that shows that sequences of ETFs generated using subsets of difference sets exist in pairs. This can be exploited to construct additional codebooks corresponding to new dimensions by taking the complement of each set of coefficients..

Proposition 5.3. *Let* $\mathbf{k}_1 \subset \mathbf{k}_2 \subset \cdots \subset \mathbf{k}_p$ *be a sequence of sets of coefficients that generate ETFs of dimensions* $d_1, \ldots, d_p$ *and codebook sizes* $N_1, \ldots, N_p$ *and* $\pi(\cdot)$ *be a permutation. Then, the sequence* $\mathbb{Z}_{N_{\pi(1)}} \setminus \mathbf{k}_{\pi(1)} \subset \mathbb{Z}_{N_{\pi(2)}} \setminus \mathbf{k}_{\pi(2)} \subset \cdots \subset \mathbb{Z}_{N_{\pi(p)}} \setminus \mathbf{k}_{\pi(p)}$ *are coefficients that generate ETFs of dimensions* $N_1 - d_1, \ldots, N_p - d_p$

5.6 Simulation Results

In this section, Monte Carlo simulations are used to demonstrate that the proposed low complexity construction performs comparably with the standard construction. While the proposed codebook construction is applicable to BSs with multiple antennas, for the purposes of simulation comparisons it is assumed that each BS has a single antenna. Also assume that each channel is Rayleigh fading and independent. The distance between BS i and the user is R_i. As such, the distribution of channel SNR between BS i and the user is given by

$$f_{\gamma_i}(\gamma) = \frac{1}{\overline{\gamma}_i} e^{-\gamma/\overline{\gamma}_i}, \tag{5.19}$$

where $\overline{\gamma}_i = P_i R_i^{-\alpha}/N_0$, $P_i = 1$ is the transmit power of BS i, $\alpha = 2$ is the path loss exponent and N_0 is the noise power. The feedback link is assumed to be error-free and zero-latency.

In Fig. 5.2, the throughput is plotted against the transmit SNR for varying numbers of BSs. The distance from each BS to the user is $R_i = 1$, $i = 1, \ldots, d$. Observe that the construction performs comparably with the exhaustive search construction for all numbers of BSs. The largest gap in performance is less than 0.15 bits/s/Hz at 15 dB and $d = 6$. Also note that the performance gain by exploiting additional BSs diminish as the number of BSs increases. This is expected as each BS is at the same distance from the user.

In Fig. 5.3, the throughput is plotted against the transmit SNR for varying numbers of BSs. In this case, the distance to the user for BS 1 is $R_1 = 100$ and $R_i = 1, 2, \ldots, d$. Observe that the construction performs comparably with the exhaustive search. In fact, the performance reduction using the proposed construction is negligible. In this figure, increasing the number of BSs results in an improved performance gain compared to the

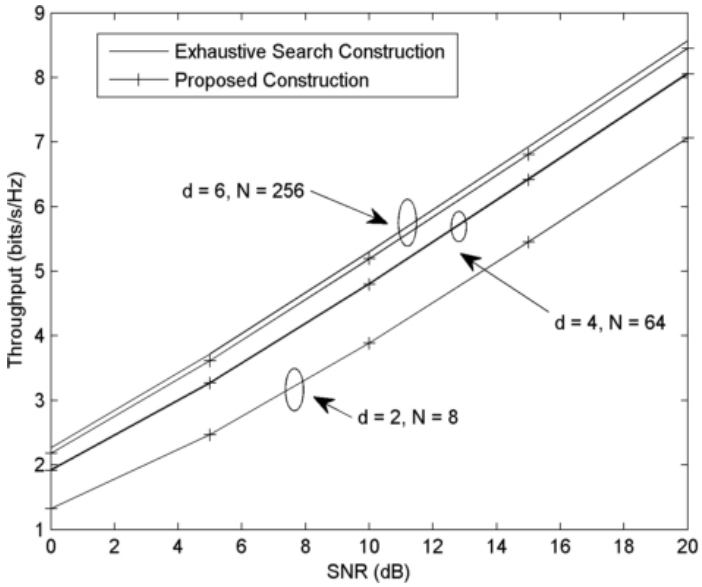


Figure 5.2: Plot of throughput versus transmit SNR when the proposed codebook and exhaustive search construction are employed for varying numbers of transmitting BSs. The distances between each BS and the user are $R_i = 1,\ i = 1, \ldots, d$.

scenario in Fig. 5.2. This is due to the fact that one BS-user link is degraded due to the large transmission distance.

5.7 Conclusion

A low complexity codebook construction was proposed for CoMP with multiple antenna BSs and a single antenna user. Our construction was based on a new recursive lower bound on the maximum inner product between codewords. This allowed us to deal with the new problem in CoMP of constructing variable dimension codebooks. A new method was also proposed to reduce storage requirements without compromising performance by requiring only one set of coefficients for several codebook dimensions. The performance of the low complexity construction was demonstrated via simulations. The construction was shown to perform comparably with the standard exhaustive search construction, while only requiring construction complexity linear in the total number of antennas and codebook size.

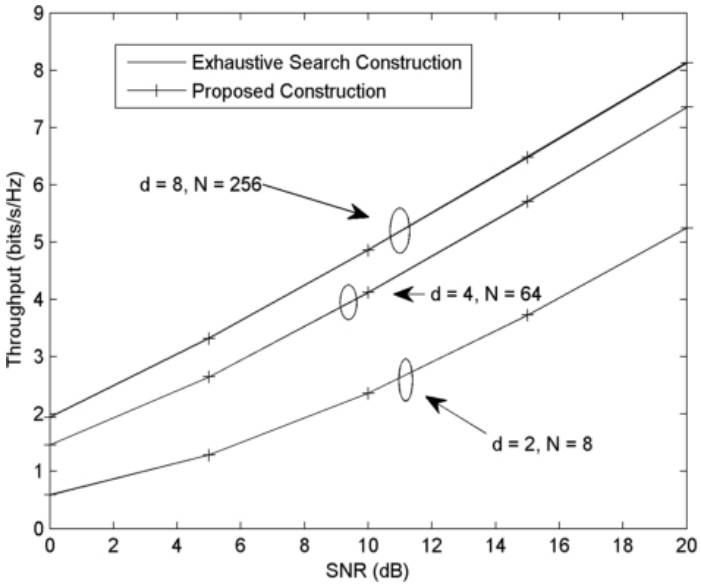


Figure 5.3: Plot of throughput versus transmit SNR when the proposed codebook and exhaustive search construction are employed for varying numbers of transmitting BSs. The distances between each BS and the user are $R_1 = 100$, $R_i = 1$, $i = 2, \ldots, d$.

Chapter 6

Resource Allocation in Multiuser MIMO for Physical Layer Security

In this chapter, resource allocation algorithms are proposed to achieve physical layer security in multiuser MIMO networks. In particular, a linear precoder based regularized channel inversion is proposed and the optimal regularization parameter derived in the large-system regime. The proposed precoder is compared with standard matched filter and zero-forcing approaches. To further improve performance, power allocation algorithms based on convex optimization techniques are developed and analyzed.

6.1 Introduction

IN current MU-MIMO systems such as LTE and 802.11n, securing transmitted data from nearby eavesdroppers is critical. In these systems, security is achieved using potentially vulnerable network layer cryptography techniques. The vulnerability is due to a reliance on the limited resources of the eavesdropper and on the unproven computational complexity of inverting the encryption algorithms [Mukherjee et al., 2010]. To enhance the protection of transmitted data and achieve perfect secrecy, methods exploiting the channel, known as physical layer security, have been proposed.

In this chapter, a linear precoder based on regularized channel inversion (RCI) [Sung et al., 2009, Joung and Lee, 2007] is proposed to achieve physical layer security in a multi-user MIMO system. In particular, the optimal regularization parameter α_{LS} and the corresponding achievable secrecy sum-rate are derived. Numerical results confirm the accuracy of the large-system analysis, even when applied to a number of users as low as 4. Moreover, the RCI precoder with α_{LS} outperforms several other linear precod-

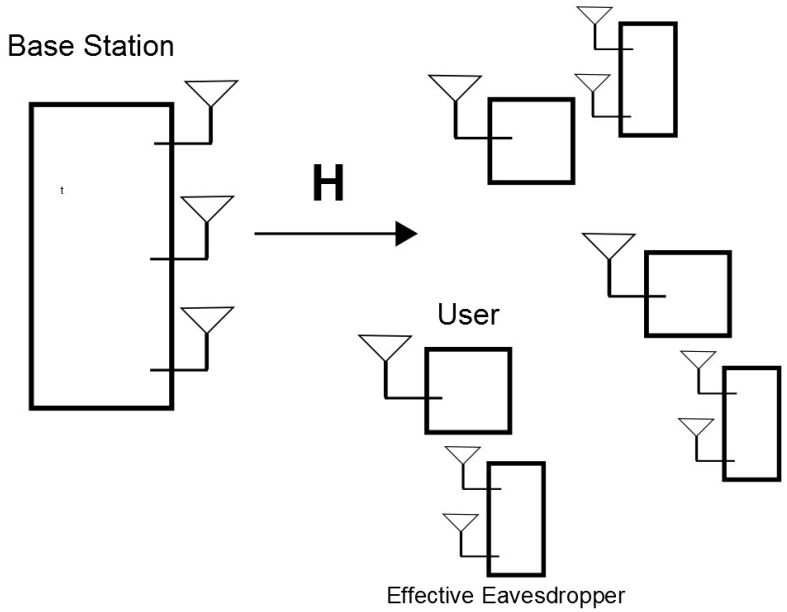


Figure 6.1: MU-MIMO network with eavesdroppers.

ing schemes. In fact, it achieves a secrecy sum-rate that has same scaling factor as the sum-rate achieved by the optimum RCI precoder without secrecy requirements.

While Chapters 3-5 were concerned with the standard performance metrics of the sum-rate and the BER, this chapter considers the secrecy sum-rate. From a mathematical point of view, this chapter can be seen to generalize the analysis of the previous chapters. In particular, the sum-rate (the focus of Chapters 4 and 5) is a special case of the secrecy sum-rate when there are no eavesdroppers. As such, standard power control algorithms based on the waterfilling solution are no longer optimal. To this end, a new iterative power control algorithm is proposed to obtain the maximum secrecy sum-rate for fixed α. The algorithm is then extended to maximize the secrecy sum-rate by jointly optimizing the regularization parameter α and the power allocation vector. The proposed power allocation algorithm outperforms RCI with α_{LS} and equal power allocation (RCI-EP) by up to 20 percent at practical values of the SNR and for 4 users and 4 transmit antennas.

6.2 System Model

Consider the downlink of a narrowband MU-MIMO system (see Fig. 6.1, consisting of a base station (BS) with M antennas which simultaneously transmits K independent confidential messages to K spatially dispersed single-antenna users. Transmission takes place over a block fading channel, where the coherence time of the channel is much longer than one symbol interval. In this model, the transmitted signal is $\mathbf{x} = [x_1, \ldots, x_M]^T \in \mathbb{C}^{M\times 1}$, and the received signal at user k is given by

$$y_k = \sum_{j=1}^{M} h_{k,j} x_j + n_k \tag{6.1}$$

where $h_{k,j} \sim \mathcal{CN}(0,1)$ is the fading gain between the j-th transmit antenna element and the k-th user, and $n_k \sim \mathcal{CN}(0,\sigma^2)$ is the noise seen at the k-th receiver. The corresponding vector equation is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{6.2}$$

where $\mathbf{H} = [h_{k,j}]$ is the $K \times M$ channel matrix, $\mathbf{y} = [y_1, \ldots, y_K]^T$ and $\mathbf{n} = [n_1, \ldots, n_K]^T$. Impose the long term power constraint $\mathrm{E}[\|\mathbf{x}\|^2] = 1$, assume that $\mathrm{E}\left[\mathbf{n}\mathbf{n}^\dagger\right] = \sigma^2\mathbf{I}$, and define the SNR $\rho = 1/\sigma^2$. The transmitted signal $\mathbf{x}$ is obtained at the BS by performing a linear processing on the confidential messages u_k, $k = 1, \ldots, K$.

It is required that the BS securely transmits each confidential message u_k, ensuring that the unintended users receive no information. This is performed at the secrecy rate $R_{s,k}$, defined as follows. Let $\Pr(\mathcal{E}_n)$ be the probability of error at the intended user, m be a confidential message, $\mathbf{y}_e^n$ be the vector of all signals received by the eavesdroppers, and $H(m|\mathbf{y}_e^n)$ be the corresponding equivocation. Then a (weak) secrecy rate $R_{s,k}$ for the intended user is achievable if there exists a sequence of $(2^{nR_{s,k}}, n)$ codes such that $\Pr(\mathcal{E}_n) \to 0$ and $\frac{1}{n}H(m|\mathbf{y}_e^n) \leq \frac{1}{n}H(m) - \varepsilon_n$ with ε_n approaching zero as $n \to \infty$ [Khisti and Wornell, 2010a].

In general, the behavior of the users cannot be determined by the BS. As a worst-case scenario, in the system under consideration it is assumed that for each intended receiver k the remaining $K-1$ users can cooperate to jointly eavesdrop on the message u_k. For

each user k, the alliance of the $K-1$ cooperating eavesdroppers is equivalent to a single eavesdropper with $K-1$ receive antennas, which is denoted by $\widetilde{k}$.

6.3 Linear Precoding

In this section, an achievable secrecy sum-rate is derived for the MU-MIMO downlink with malicious users by using a linear precoder. Although suboptimal, linear precoding schemes are of particular interest because of their low-complexity implementations and because they can control the amount of crosstalk between the users [Yoo and Goldsmith, 2006, Spencer et al., 2004, Peel et al., 2005, Joham et al., 2005]. Next, secrecy sum-rate achievable by the RCI precoder is obtained. RCI is a linear precoding scheme that was proposed to serve multiple users in the MU-MIMO downlink channel, which has better performance than plain channel inversion, especially at low SNR [Peel et al., 2005].

6.3.1 Preliminaries

In linear precoding, the transmitted vector $\mathbf{x}$ is derived from the vector containing the confidential messages $\mathbf{u} = [u_1, \ldots, u_K]^T$ through a deterministic linear transformation (precoding) [Yoo and Goldsmith, 2006, Spencer et al., 2004, Peel et al., 2005, Joham et al., 2005]. Assume that the entries of $\mathbf{u}$ are chosen independently, satisfying $\mathrm{E}[|u_k|^2] = 1$. Assume spatially homogeneous users, i.e. each user experiences the same received signal power on average, thus the model assumes that their distances from the transmitter are similar.

Let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ be the $M \times K$ precoding matrix, where $\mathbf{w}_k$ is the k-th column of $\mathbf{W}$. Then the transmitted signal and the power constraint are, respectively:

$$\mathbf{x} = \frac{1}{\sqrt{\gamma}}\mathbf{W}\mathbf{u} = \frac{1}{\sqrt{\gamma}}\sum_{k=1}^{K}\mathbf{w}_k u_k, \tag{6.3}$$

$$\mathrm{E}\left[\|\mathbf{x}\|^2\right] = \frac{1}{\gamma}\mathrm{E}\left[\|\mathbf{W}\mathbf{u}\|^2\right] = \frac{1}{\gamma}\sum_{k=1}^{K}\|\mathbf{w}_k\|^2 = 1, \tag{6.4}$$

where $\gamma = \mathrm{tr}\{\mathbf{W}^\dagger\mathbf{W}\}$ is the long-term power normalization constant.

The two key linear precoders considered in this chapter are zero-forcing and regularized channel inversion. In particular, the zero-forcing precoder is given by

$$\mathbf{W}_{ZF} = \mathbf{H}^{\dagger}\mathbf{H}\mathbf{H}^{\dagger}. \tag{6.5}$$

On the other hand, the regularized channel inversion precoder is given by

$$\mathbf{W}_{RCI} = \mathbf{H}^{\dagger}\left(\mathbf{H}\mathbf{H}^{\dagger} + \alpha\mathbf{I}_K\right)^{-1}. \tag{6.6}$$

Observe that the zero-forcing precoder is a special case of the regularized channel inversion precoder when $\alpha = 0$.

6.3.2 Achievable Secrecy Sum-Rates with Linear Precoding

By employing the linear precoding in (6.3), the signals observed at receivers k and $\widetilde{k}$ are, respectively

$$\begin{aligned} y_k &= \frac{1}{\sqrt{\gamma}}\mathbf{h}_k^{\dagger}\mathbf{w}_k u_k + \frac{1}{\sqrt{\gamma}}\sum_{j\neq k}\mathbf{h}_k^{\dagger}\mathbf{w}_j u_j + n_k \\ \mathbf{y}_{\widetilde{k}} &= \frac{1}{\sqrt{\gamma}}\sum_{k}\mathbf{H}_{\widetilde{k}}\mathbf{w}_k u_k + \mathbf{n}_{\widetilde{k}} \end{aligned} \tag{6.7}$$

where $\mathbf{n}_{\widetilde{k}} = [n_1, \ldots, n_{k-1}, n_{k+1}, \ldots, n_K]^T$, $\mathbf{h}_k^{\dagger}$ is the k-th row of $\mathbf{H}$, and $\mathbf{H}_{\widetilde{k}}$ is a matrix obtained from $\mathbf{H}$ by eliminating the k-th row. The channel in (6.7) is a multi-input, single-output, multi-eavesdropper (MISOME) wiretap channel [Khisti and Wornell, 2010a]. The transmitter, the intended receiver and the eavesdropper of this MISOME wiretap channel are equipped with M, 1 and $K-1$ virtual antennas, respectively. Due to the simultaneous transmission of the K messages, user k experiences noise and interference from all the u_j, $j \neq k$.

In the following, an achievable secrecy sum-rate R_s is derived for the MU-MIMO system with malicious users. Although the design of codes for the MU-MIMO channel with security constraints is not the focus of this chapter, the achievability of R_s is proved with a code construction based on independent codebooks and linear precoding.

Lemma 6.1 (Codebook construction)**.** *An achievable secrecy sum-rate R_s for the MU-MIMO*

system with malicious users is given by

$$R_s \triangleq \sum_{k=1}^{K} R_{s,k}, \tag{6.8}$$

where $R_{s,k}$ is an achievable secrecy rate for the k-th MISOME wiretap channel (6.7), $k = 1, \ldots, K$.

Proof: Assume that the BS uses independent codebooks for each user, where each codebook is a code for the scalar wiretap channel [Khisti and Wornell, 2010a]. The confidential message u_k is obtained as a codeword independently drawn from the code $\mathcal{C}_k$, corresponding to the k-th user. The rate $R_{s,k}$ of the code $\mathcal{C}_k$ is chosen according to the secrecy rate achievable for user k in the presence of eavesdropper $\widetilde{k}$, i.e. by the secrecy rate achievable for the MISOME wiretap channel (6.7). The existence of such code is guaranteed by the definition of secrecy rate [Csiszár and Körner, 1978]. To construct the vector codeword for the broadcast channel, the scalar codewords for each MISOME wiretap channel are stacked according to $\mathbf{u} = [u_1, \ldots, u_K]^T$, and no additional binning is required. The vector $\mathbf{u}$ is then linearly precoded as in (6.3), which means that each message u_k is transmitted by beamforming (i.e. signaling with rank one covariance) along the direction $\mathbf{w}_k$. The secrecy sum-rate R_s is then by definition the sum of the simultaneously achievable secrecy rates $R_{s,k}$. ■

Lemma 6.2. *An achievable secrecy rate $R_{s,k}$ for the MISOME wiretap channel (6.7) is given by*

$$R_{s,k} = \left[\log_2\left(1 + \mathrm{SINR}_k\right) - \log_2\left(1 + \mathrm{SINR}_{\widetilde{k}}\right)\right]^+, \tag{6.9}$$

where SINR_k and $\mathrm{SINR}_{\widetilde{k}}$ are the signal-to-interference-plus-noise ratios for the message u_k at the intended receiver k and the eavesdropper $\widetilde{k}$, respectively.

Proof: By noting that the MISOME wiretap channel (6.7) is a nondegraded broadcast channel [Khisti and Wornell, 2010a], the secrecy capacity is given by [Csiszár and Körner, 1978]:

$$C_s = \max_{u_k \to \mathbf{w}_k u_k \to y_k, \mathbf{y}_{\widetilde{k}}} I\left(u_k; y_k\right) - I\left(u_k; \mathbf{y}_{\widetilde{k}}\right) \tag{6.10}$$

where $I(x;y)$ denotes mutual information between two random variables x and y. The secrecy capacity C_s is given by the difference of the mutual informations at the intended user and at the eavesdropper, respectively. C_s is achieved by maximizing over all joint probability distributions such that a Markov chain $u_k \rightarrow \mathbf{w}_k u_k \rightarrow y_k, \mathbf{y}_{\widetilde{k}}$ is formed, where u_k is an auxiliary input variable. By evaluating (6.10) with $u_k \sim \mathcal{CN}(0,1)$ and with the linearly precoded data $\mathbf{w}_k u_k$, an achievable secrecy rate $R_{s,k}$ is obtained for the MISOME wiretap channel (6.7) given by

$$R_{s,k} = \left[I\Big(u_k; y_k\Big) - I\Big(u_k; \mathbf{y}_{\widetilde{k}}\Big) \right]^+ \tag{6.11}$$

$$\stackrel{a}{=} \left[I\Big(\mathbf{w}_k u_k; y_k\Big) - I\Big(\mathbf{w}_k u_k; \mathbf{y}_{\widetilde{k}}\Big) \right]^+, \tag{6.12}$$

where (a) follows from $\mathbf{w}_k u_k$ being a deterministic function of u_k [Khisti and Wornell, 2010a]. Equation (6.9) then follows from (6.12) and from the statistics of u_k. ■

From equation (6.9) it is clearly observed that for high-performance linear precoder design an efficient tradeoff between maximizing SINR_k and minimizing $\mathrm{SINR}_{\widetilde{k}}$ is required.

Theorem 6.1. *A secrecy sum-rate achievable by MU-MIMO linear precoding is given by*

$$R_s = \sum_{k=1}^{K} \left[\log_2 \frac{1 + \frac{\left|\mathbf{h}_k^\dagger \mathbf{w}_k\right|^2}{\gamma\sigma^2 + \sum_{j\neq k}\left|\mathbf{h}_k^\dagger \mathbf{w}_j\right|^2}}{1 + \frac{\left\|\mathbf{H}_{\widetilde{k}} \mathbf{w}_k\right\|^2}{\gamma\sigma^2}} \right]^+ . \tag{6.13}$$

Proof: By using Lemma 6.1 and Lemma 6.2, an achievable secrecy sum-rate is obtained as the sum of the secrecy rates $R_{s,k}$ in (6.9). A lower bound on the quantities $R_{s,k}$ can be obtained by considering a genie-aided eavesdropper which observes not only the signals $\mathbf{y}_{\widetilde{k}}$ received by its $K-1$ antennas, but also all the confidential messages $u_j, j \neq k$. Such channel clearly has an achievable secrecy rate smaller than the original channel. The genie-aided eavesdropper $\widetilde{k}$ can perform interference cancellation, and it does not see any undesired signal term apart from the received noise $\mathbf{n}_{\widetilde{k}}$.

According to the previous considerations, the signals at the intended receiver and the

eavesdropper of the k-th equivalent MISOME wiretap channel become, respectively:

$$\begin{aligned} y_k &= \frac{1}{\sqrt{\gamma}}\mathbf{h}_k^\dagger \mathbf{w}_k u_k + \frac{1}{\sqrt{\gamma}}\sum_{j\neq k}\mathbf{h}_k^\dagger \mathbf{w}_j u_j + n_k \\ \mathbf{y}_{\widetilde{k}} &= \frac{1}{\sqrt{\gamma}}\mathbf{H}_{\widetilde{k}}\mathbf{w}_k u_k + \mathbf{n}_{\widetilde{k}} \end{aligned} \tag{6.14}$$

For the k-th equivalent MISOME wiretap channel in (6.14), the SINRs at the intended user and the eavesdropper are, respectively:

$$\mathrm{SINR}_k = \frac{\left|\mathbf{h}_k^\dagger \mathbf{w}_k\right|^2}{\gamma\sigma^2 + \sum_{j\neq k}\left|\mathbf{h}_k^\dagger \mathbf{w}_j\right|^2}, \tag{6.15}$$

$$\mathrm{SINR}_{\widetilde{k}} = \frac{\left\|\mathbf{H}_{\widetilde{k}}\mathbf{w}_k\right\|^2}{\gamma\sigma^2}. \tag{6.16}$$

Since the noise in $\mathbf{y}_{\widetilde{k}}$ in (6.14) is spatially white, the optimal receive filter at $\widetilde{k}$ is the matched filter $(\mathbf{H}_{\widetilde{k}}\mathbf{w}_k)^\dagger$. Equation (6.16) then follows. For a given channel $\mathbf{H}$, substituting (6.15) and (6.16) into (6.9) and then into (6.8) yields (6.13). ■

For the remainder of the chapter, equation (6.13) is referred to as the secrecy sum-rate. Note that it depends on the choice of the precoding matrix $\mathbf{W}$, as well as on the channel $\mathbf{H}$ and the noise variance σ^2. A possible choice for $\mathbf{W}$, based on regularized channel inversion, is discussed in the following.

6.3.3 Achievable Secrecy Sum-Rates with Regularized Channel Inversion

Now consider RCI precoding for the MU-MIMO downlink with malicious users. Although CI precoding can achieve secrecy by canceling all signals leaked at the unintended users, this comes at the cost of a poor sum-rate. The RCI precoder has better performance than plain CI, particularly at low SNR [Peel et al., 2005]. For each message u_k, RCI precoding achieves a tradeoff between the signal power at the k-th intended user and the crosstalk at the other $(K-1)$ unintended users for each signal. The crosstalk causes interference to the unintended users. In the case when the unintended users are acting maliciously, the crosstalk also causes information leakage. Therefore, RCI achieves

a tradeoff between signal power, interference, and information leakage.

With RCI precoding, linear processing exploiting regularization is applied to the vector of messages $\mathbf{u}$ [Peel et al., 2005]. The RCI precoding matrix is given by

$$\mathbf{W} = \mathbf{H}^\dagger \left(\mathbf{H}\mathbf{H}^\dagger + \alpha \mathbf{I}_K \right)^{-1}. \tag{6.17}$$

The transmitted signal $\mathbf{x}$ after RCI precoding can be written as

$$\mathbf{x} = \frac{1}{\sqrt{\gamma}} \mathbf{W}\mathbf{u} = \frac{1}{\sqrt{\gamma}} \mathbf{H}^\dagger \left(\mathbf{H}\mathbf{H}^\dagger + \alpha \mathbf{I}_K \right)^{-1} \mathbf{u} \tag{6.18}$$

$$= \frac{1}{\sqrt{\gamma}} \left(\mathbf{H}^\dagger \mathbf{H} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{H}^\dagger \mathbf{u}. \tag{6.19}$$

The latter passes through the channel, producing the vector of received signals

$$\mathbf{y} = \frac{1}{\sqrt{\gamma}} \mathbf{H} \left(\mathbf{H}^\dagger \mathbf{H} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{H}^\dagger \mathbf{u} + \mathbf{n}. \tag{6.20}$$

The function of the real nonnegative regularization parameter α is to improve the behavior of the inverse, although it also produces non-zero crosstalk terms in (6.20).

Using RCI precoding, the SINRs (6.15) and (6.16) at the intended user k and the eavesdropper $\widetilde{k}$ become

$$\mathrm{SINR}_k = \frac{\left| \mathbf{h}_k^\dagger \left(\mathbf{H}^\dagger \mathbf{H} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{h}_k \right|^2}{\gamma \sigma^2 + \sum_{j \neq k} \left| \mathbf{h}_k^\dagger \left(\mathbf{H}^\dagger \mathbf{H} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{h}_j \right|^2}, \tag{6.21}$$

$$\mathrm{SINR}_{\widetilde{k}} = \frac{\left\| \mathbf{H}_{\widetilde{k}} \left(\mathbf{H}^\dagger \mathbf{H} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{h}_k \right\|^2}{\gamma \sigma^2}, \tag{6.22}$$

where

$$\gamma = \mathrm{tr} \left\{ \mathbf{H}^\dagger \mathbf{H} \left(\mathbf{H}^\dagger \mathbf{H} + \alpha \mathbf{I}_K \right)^{-2} \right\}. \tag{6.23}$$

To simplify (6.21) and (6.22), the following quantities are introduced

$$A_k = \mathbf{h}_k^\dagger \left(\mathbf{H}_{\widetilde{k}}^\dagger \mathbf{H}_{\widetilde{k}} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{h}_k \quad \text{and} \tag{6.24}$$

$$B_k = \mathbf{h}_k^\dagger \left(\mathbf{H}_{\widetilde{k}}^\dagger \mathbf{H}_{\widetilde{k}} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{H}_{\widetilde{k}}^\dagger \mathbf{H}_{\widetilde{k}} \left(\mathbf{H}_{\widetilde{k}}^\dagger \mathbf{H}_{\widetilde{k}} + \alpha \mathbf{I}_K \right)^{-1} \mathbf{h}_k. \tag{6.25}$$

It is then possible to express (6.21) as [Nguyen et al., 2009]

$$\mathrm{SINR}_k = \frac{A_k^2}{B_k + \gamma\sigma^2 \left(1 + A_k\right)^2}. \tag{6.26}$$

In a similar fashion, (6.22) is rewritten as

$$\mathrm{SINR}_{\widetilde{k}} = \frac{B_k}{\gamma\sigma^2 \left(1 + A_k\right)^2}. \tag{6.27}$$

By substituting (6.26) and (6.27) into (6.9) and then into (6.8) the following expression is obtained for the secrecy sum-rates achievable with RCI precoding

$$R_s = \sum_{k=1}^{K} \left[\log_2 \frac{1 + \frac{A_k^2}{B_k + \gamma\sigma^2 (1+A_k)^2}}{1 + \frac{B_k}{\gamma\sigma^2 (1+A_k)^2}} \right]^+. \tag{6.28}$$

6.4 Large-System Analysis

In this section, the performance of the RCI precoder is analyzed in the large-system regime, where both the number of transmit antennas M and the number of receivers K approach infinity in a fixed ratio. Closed-form expressions are derived for the optimal regularization parameter and the optimal secrecy sum-rate achievable with RCI in the large-system regime. The secrecy sum-rate achieved by the optimized RCI precoder is then compared with several other linear precoding schemes. Finally, the sum-rate loss due to the secrecy requirements is evaluated. In this section, the key focus is on the case $K = M$ due to its practical importance and the potential for high performance [Peel et al., 2005].

6.4.1 Secrecy Sum-Rates in the Large-System Regime

Define $\xi = \alpha / K$ as the normalized regularization parameter, and note that as $K \to \infty$, the quantities (6.23), (6.24) and (6.25) converge (almost surely) to [Nguyen et al., 2009]

$$\lim_{K\to\infty} \gamma = g(\xi) + \xi \frac{d}{d\xi} g(\xi), \tag{6.29}$$

$$\lim_{K\to\infty} A_k = g(\xi), \tag{6.30}$$

$$\lim_{K\to\infty} B_k = g(\xi) + \xi \frac{d}{d\xi} g(\xi), \tag{6.31}$$

respectively, where

$$g(\xi) = \frac{1}{2}\sqrt{1 + \frac{4}{\xi}} - \frac{1}{2}. \tag{6.32}$$

By substituting the above expressions in (6.26) and (6.27), one can conclude that as $K \to \infty$, all the SINRs at the intended user k and at the eavesdropper $\widetilde{k}$ converge to a non-random function of the parameter ξ and the noise variance σ^2. Moreover, these quantities are the same for all confidential messages u_k, as $K \to \infty$. Hence, in the large-system regime, it is possible to write the secrecy sum-rate with RCI precoder as

$$R_s \simeq R_{s,\infty} \triangleq K \left[\log_2 \frac{1 + \frac{\rho g(\xi)^2}{\left[\rho + (1+g(\xi))^2\right]\left[g(\xi) + \xi \frac{d}{d\xi} g(\xi)\right]}}{1 + \frac{\rho}{(1+g(\xi))^2}} \right]^+$$
$$\text{as } K \to \infty. \tag{6.33}$$

6.4.2 Selection of the Optimal Regularization Parameter

The value of the asymptotic secrecy sum-rate $R_{s,\infty}$ in (6.33) depends on the normalized regularization parameter ξ. The optimal value ξ_{opt} is now derived that maximizes $R_{s,\infty}$.

Lemma 6.3. *The optimal normalized regularization parameter in the large-system regime is given by*

$$\xi_{\text{opt}} = \frac{1}{3\rho + 1 + \sqrt{3\rho + 1}}. \tag{6.34}$$

Proof: The value of ξ_{opt} is obtained as the stationary point of the secrecy sum-rate $R_{s,\infty}$, which can be found by setting to zero the derivative of the logarithm in (6.33), by applying some algebraic manipulations, and showing that the maximum value is non-negative. ■

As in the case with no secrecy requirements, the value of ξ_{opt} is a function of the SNR ρ. In a multiuser channel without secrecy requirements, the choice $\xi = 1/\rho$ is optimal for large K, as it maximizes the sum-rate of the system [Peel et al., 2005]. However, this value is no longer optimal in a multiuser channel with malicious users. In fact, because of the secrecy requirements the crosstalk terms appear twice in the secrecy sum-rate expression (6.13). As a consequence, $\xi = 1/\rho$ is too large and gives too much crosstalk to the other users. This was proven in [Geraci et al., 2011] and it is also easily confirmed by the following inequality:

$$\xi_{\text{opt}} < \frac{1}{3\rho} < \frac{1}{\rho} \quad \forall \rho. \tag{6.35}$$

Similarly to the case with no secrecy requirements, ξ_{opt} decreases as ρ is increased. The high-SNR asymptote of ξ_{opt} is given by

$$\xi_{\text{opt}} \simeq \frac{1}{3\rho}, \quad \text{as } \rho \to \infty \tag{6.36}$$

and ξ_{opt} tends to zero if $\rho \to \infty$.

Unlike the case with no confidentiality, the optimum normalized regularization parameter is upper bounded and it does not tend to infinity as ρ tends to zero. The low-SNR asymptote of ξ_{opt} is

$$\xi_{\text{opt}} = \frac{1}{2}, \quad \text{for } \rho = 0. \tag{6.37}$$

In the remainder of the chapter, the unnormalized large-system regularization parameter is denoted by $\alpha_{\text{LS}} = K\xi_{\text{opt}}$.

6.4.3 Optimal Secrecy Sum-Rate

It is now possible to obtain an expression for the optimal secrecy sum-rate of the RCI precoder in the large-system regime. The optimal secrecy sum-rate is a function of the SNR ρ and the number of users K only.

Theorem 6.2. *The optimal secrecy sum-rate $R_{s,\infty}^{\mathrm{RCI}}$ achievable by the RCI precoder in the large-system regime is given by*

$$R_{s,\infty}^{\mathrm{RCI}} \triangleq \max_{\xi} R_{s,\infty} = K \log_2 \frac{9\rho + 2 + (6\rho+2)\sqrt{3\rho+1}}{4\,(4\rho+1)}. \tag{6.38}$$

Proof: Equation (6.38) is obtained by substituting (6.34) in (6.33) and applying some algebraic manipulations. ■

The secrecy sum-rate $R_{s,\infty}^{\mathrm{RCI}}$ in (6.38) satisfies

$$R_{s,\infty}^{\mathrm{RCI}} > 0 \quad \forall \rho > 0, \tag{6.39}$$

and the high-SNR asymptote of $R_{s,\infty}^{\mathrm{RCI}}$ is given by

$$R_{s,\infty}^{\mathrm{RCI}} \simeq \frac{K}{2} \log_2 \frac{27}{64} + \frac{K}{2} \log_2 \rho, \quad \text{as } \rho \to \infty. \tag{6.40}$$

Therefore, in the large-system regime the secrecy sum-rate for optimal ξ scales logarithmically with high SNR, and it scales linearly as $K/2$ with the number of users.

Although for $K \to \infty$ the number of eavesdroppers $K-1$ for each message tends to infinity, a positive secrecy sum-rate is still achievable. This occurs because the number of transmit antennas $M = K$ also tends to infinity, and it is larger than the number of eavesdroppers. Therefore, for each message the transmitter is able to control the amount of interference and information leakage.

Now the per-user secrecy rate achieved by RCI is compared with the secrecy capacity of the MISOME channel, C_s^{MISOME}, in the high-SNR regime. The former is obtained by dividing (6.40) by the number of users K, and it can be further approximated by

$$\frac{R_{s,\infty}^{\mathrm{RCI}}}{K} \simeq \frac{1}{2} \log_2 \rho, \quad \text{as } \rho \to \infty. \tag{6.41}$$

The value of C_s^{MISOME} was obtained in [Khisti and Wornell, 2010a], and it can be approx-

imated by the following lower bound

$$C_s^{\text{MISOME}} \geq \frac{1}{2}\log_2 \rho, \quad \text{as } \rho \to \infty, \tag{6.42}$$

which is tight at high SNR [Khisti and Wornell, 2010a]. Note that in C_s^{MISOME} from [Khisti and Wornell, 2010a] a single-user system is considered. Therefore, only one message is transmitted to one legitimate user, and the user does not experience any interference. For large SNR, the RCI precoder achieves a per-user secrecy rate which is the same as the secrecy capacity of a single-user system.

6.4.4 Comparison to Other Linear Schemes

In the following, the secrecy sum-rate in (6.38) achieved by the RCI precoder is compared with ξ_{opt} to the secrecy sum-rates obtained from (6.33) by using: 1) $\xi = 0$ (CI precoder), 2) $\xi \to \infty$ (matched-filter precoder) and 3) $\xi = 1/\rho$ (optimum RCI precoder without secrecy requirements).

The aim of the CI precoder is to cancel all the interference and information leakage, therefore yielding to a secrecy sum-rate that coincides with the sum-rate. Note that for the CI precoder it is $\xi = 0$, and the precoding matrix is given by

$$\mathbf{W} = \mathbf{H}^{\dagger}\left(\mathbf{H}\mathbf{H}^{\dagger}\right)^{-1}. \tag{6.43}$$

In order for the inverse in (6.43) to exist, it is required that $K \leq M$.

The secrecy sum-rate achieved by CI in the large-system regime grows at most sublinearly with $K \to \infty$. In fact,

$$\lim_{\xi \to 0} \lim_{K \to \infty} \frac{R_s}{K} = 0. \tag{6.44}$$

This result is consistent with [Nguyen et al., 2009], where it was shown that the CI precoder performs poorly in the large-system regime when the number of antennas equals the number of users.

Similarly, the secrecy sum-rate achieved when $\xi \to \infty$ (matched-filter precoding) is calculated. Here, the transmitter beamforms in a direction such as to maximize the sig-

nal strenght of each user, without taking into account the interference it creates and the amount of resulting information leakage. The secrecy sum-rate achieved by matched-filter precoding in the large-system regime is zero. In fact,

$$\lim_{\xi\to\infty}\lim_{K\to\infty}\frac{R_s}{K}=\left[\log_2\frac{2\rho+1}{(\rho+1)^2}\right]^+=0. \tag{6.45}$$

Clearly, matched-filter precoding performs poorly compared to the optimal RCI precoder. This is due to the intended user suffering from a large amount of interference, while the eavesdroppers may cancel the interference by cooperating.

Finally, consider $\xi=1/\rho$, which is the value that maximizes the sum-rate of the system without secrecy requirements [Peel et al., 2005]. The secrecy sum-rate $R^{\circ}_{s,\infty}$ achieved by RCI with $\xi=1/\rho$ in the large-system regime is given by

$$R^{\circ}_{s,\infty}=K\log_2\frac{4\rho+1+(2\rho+1)\sqrt{4\rho+1}}{2\,(4\rho+1)}. \tag{6.46}$$

Observe that the RCI scheme with $\xi=1/\rho$ outperforms the CI and the matched-filter precoding schemes in the large-system regime, but it is suboptimal compared to the use of ξ_{opt}. For high SNR, the per-antenna secrecy sum-rate gain provided by using $\xi=\xi_{\text{opt}}$ in place of $\xi=1/\rho$ is given by

$$\lim_{\rho\to\infty}\frac{R^{\text{RCI}}_{s,\infty}-R^{\circ}_{s,\infty}}{K}=\log_2\frac{3\sqrt{3}}{4}\approx 0.38 \text{ bits}. \tag{6.47}$$

6.4.5 Secrecy Loss

Now the *secrecy loss* is considered; i.e. the sum-rate loss due to the secrecy requirements. Define this as the difference between the optimal sum-rate R°_{∞} without secrecy requirements and the secrecy sum-rate $R^{\text{RCI}}_{s,\infty}$ in (6.38). The sum-rate R°_{∞} is obtained with RCI and $\xi=1/\rho$, and it is given by [Nguyen et al., 2009]

$$R^{\circ}_{\infty}=K\log_2\frac{1+\sqrt{4\rho+1}}{2}. \tag{6.48}$$

The high-SNR asymptote of the sum-rate in (6.48) is

$$R_{\infty}^{\circ} \simeq \frac{K}{2} \log_2 \rho, \quad \text{as } \rho \to \infty. \tag{6.49}$$

For high SNR, the per-antenna secrecy loss is given by

$$\lim_{\rho \to \infty} \frac{R_{\infty}^{\circ} - R_{s,\infty}^{\mathrm{RCI}}}{K} = \frac{1}{2} \log_2 \frac{64}{27} \approx 0.62 \text{ bits}. \tag{6.50}$$

By comparing (6.49) to (6.40), one can conclude that the secrecy requirements do not change the linear scaling factor for large SNR. In other words, the RCI precoder with ξ_{opt} achieves a secrecy sum-rate that has same scaling factor $K/2$ as the sum-rate achieved by the optimum RCI precoder without secrecy requirements in [Peel et al., 2005]. The RCI precoder with ξ_{opt} can achieve secrecy with a penalty in terms of the per-antenna sum-rate given by (6.50). The secrecy loss (6.50) corresponds to a power loss of a factor $64/27 \approx 3.75$dB. Therefore, the RCI precoder with ξ_{opt} can achieve secrecy without reducing the sum-rate of the system, as long as the transmitted power is increased by 3.75dB.

6.5 Power Allocation

In this section, power allocation for the RCI precoder is considered. First, a new algorithm is proposed to obtain the power allocation vector $\mathbf{p}$ which achieves the optimal secrecy sum-rate with a fixed regularization parameter α. The algorithm is then extended to jointly optimize $\mathbf{p}$ and α.

6.5.1 Achievable Secrecy Sum-Rates

Consider the RCI precoding matrix with arbitrary power allocation given by

$$\mathbf{W}_{\mathrm{p}} = \mathbf{W}\mathbf{D} = \mathbf{H}^{\dagger}(\mathbf{H}\mathbf{H}^{\dagger} + \alpha \mathbf{I})^{-1}\mathbf{D}, \tag{6.51}$$

Table 6.1: Proposed algorithms for power allocation.

Algorithm 1	**Algorithm 2**
Initialize iteration counter $t = 0$ Initialize all $a_k^{(t)} = 1$, $b_k^{(t)} = 0$ **repeat** Solve (6.60) to obtain $\widetilde{\mathbf{p}}^{(t)}$ Update $a_k^{(t)}$, $b_k^{(t)}$ at $z_0 = \text{SINR}_k(\widetilde{\mathbf{p}}^{(t)})$ Increment t **until** convergence Obtain $p_k = e^{\widetilde{p}_k}$, $k = 1, \ldots, K$	Initialize iteration counter $t_1 = 0$, $t_2 = 0$ Initialize $p_k = 1/\gamma$, and set $\widetilde{p}_k = \log p_k$, $k = 1, \ldots, K$ Initialize $\alpha_0 = K\xi_{\text{opt}}$ using equation (6.34) **repeat** Increment t_1 Obtain $\alpha_{t_1}^*$ using steepest descent with α_{t_1-1} as initial point Initialize all $a_k^{(t_2)} = 1$, $b_k^{(t_2)} = 0$ **repeat** Solve (6.60) to obtain $\widetilde{\mathbf{p}}^{(t_2)}$ Update $a_k^{(t_2)}$, $b_k^{(t_2)}$ at $z_0 = \text{SINR}_k(\alpha_{t_1}^*, \widetilde{\mathbf{p}}^{(t_2)})$ Increment t_2 **until** convergence Set $\widetilde{\mathbf{p}} = \widetilde{\mathbf{p}}^{(t_2)}$ **until** convergence Obtain $p_k = e^{\widetilde{p}_k}$, $k = 1, \ldots, K$

where $\mathbf{D} = \text{diag}(\sqrt{\mathbf{p}})$, and $\mathbf{p} = [p_1, \ldots, p_K]^T$ is the power allocation vector. The vector $\mathbf{p}$ must be chosen such that the power constraint $\text{tr}\left\{\mathbf{W}_{\text{p}}^{\dagger}\mathbf{W}_{\text{p}}\right\} = 1$ is met. Clearly, (6.51) generalizes the RCI precoder $\mathbf{W}$ with equal power allocation (RCI-EP) in (6.17).

When the precoder $\mathbf{W}_{\text{p}}$ is used, the SINR at the k-th intended user, given by (6.15), becomes

$$\text{SINR}_k = \frac{p_k |\mathbf{h}_k^{\dagger}\mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^{\dagger}\mathbf{w}_j|^2 + \sigma^2}, \tag{6.52}$$

and the SINR at the eavesdropper $\widetilde{k}$, given by (6.16), becomes

$$\mathrm{SINR}_{\widetilde{k}} = \frac{p_k \|\mathbf{H}_{\widetilde{k}} \mathbf{w}_k\|^2}{\sigma^2} = \frac{p_k \sum_{j \neq k} |\mathbf{h}_j^\dagger \mathbf{w}_k|^2}{\sigma^2}. \tag{6.53}$$

From (6.52) and (6.53), the achievable secrecy sum-rate is obtained with power allocation

$$R_{\mathrm{s}}^{\mathrm{PA}} = \sum_{k=1}^{K} \left[\log_2 \frac{1 + \frac{p_k |\mathbf{h}_k^\dagger \mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^\dagger \mathbf{w}_j|^2 + \sigma^2}}{1 + \frac{p_k \sum_{j \neq k} |\mathbf{h}_j^\dagger \mathbf{w}_k|^2}{\sigma^2}} \right]^+. \tag{6.54}$$

6.5.2 Power Control

To obtain the optimal power allocation vector $\mathbf{p}$, the following non-convex optimization problem is required to be solved

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & R_s^{\mathrm{PA}}(\mathbf{p}) \\ \text{subject to} \quad & \mathrm{tr}\left\{ \mathbf{W}_{\mathrm{p}}^\dagger \mathbf{W}_{\mathrm{p}} \right\} \leq 1, \end{aligned} \tag{6.55}$$

where $R_s^{\mathrm{PA}}(\mathbf{p})$ is given by (6.54), $\mathbf{W}_{\mathrm{p}}$ is given by (6.51), and the maximum total transmit power over all antennas is one. In the following, the notation $[\cdot]^+$ in (6.54) will be ignored in the maximization problem. In fact, any negative term in the sum can be replaced by zero (thus increasing the sum) by using $p_k = 0$ which is always feasible.

The problem (6.55) is then reformulated by applying the transformation $\widetilde{p}_k = \log p_k$, $k = 1 \ldots, K$, and obtain the optimization problem

$$\begin{aligned} \underset{\widetilde{\mathbf{p}}}{\text{maximize}} \quad & R_s^{\mathrm{PA}}(\widetilde{\mathbf{p}}) \\ \text{subject to} \quad & \mathrm{tr}\left\{ \mathbf{W}_{\mathrm{p}}^\dagger \mathbf{W}_{\mathrm{p}} \right\} \leq 1, \end{aligned} \tag{6.56}$$

where $\widetilde{\mathbf{p}} = [\widetilde{p}_1, \ldots, \widetilde{p}_K]^T$.

Lemma 6.4. *The second term of the objective function, $R_{\mathrm{s}}^{\mathrm{PA}}(\widetilde{\mathbf{p}})$, of (6.56) is concave.*

Proof: The second term and its first and second derivatives are

$$\begin{aligned} -\log_2\left(1+\mathrm{SINR}_{\widetilde{k}}\right) &= -\log_2\left(1+\frac{e^{\widetilde{p}_k}\sum_{j\neq k}|\mathbf{h}_j^\dagger\mathbf{w}_k|^2}{\sigma^2}\right), \\ \frac{-\partial\log_2\left(1+\mathrm{SINR}_{\widetilde{k}}\right)}{\partial\widetilde{p}_k} &= \frac{-(\log_2 e)e^{\widetilde{p}_k}\sum_{j\neq k}|\mathbf{h}_j^\dagger\mathbf{w}_k|^2}{\sigma^2+e^{\widetilde{p}_k}\sum_{j\neq k}|\mathbf{h}_j^\dagger\mathbf{w}_k|^2}, \\ \frac{-\partial^2\log_2\left(1+\mathrm{SINR}_{\widetilde{k}}\right)}{\partial\widetilde{p}_k^2} &= \frac{-(\log_2 e)e^{\widetilde{p}_k}\sum_{j\neq k}|\mathbf{h}_j^\dagger\mathbf{w}_k|^2\sigma^2}{\left(\sigma^2+e^{\widetilde{p}_k}\sum_{j\neq k}|\mathbf{h}_j^\dagger\mathbf{w}_k|^2\right)^2}\leq 0. \end{aligned} \tag{6.57}$$

Hence, by the second order condition [30, §3.4.3], $-\log_2\left(1+\mathrm{SINR}_{\widetilde{k}}\right)$ is concave. ■

In order to solve the problem (6.56), a modified version of the method in [Papandriopoulos et al., 2008] and [Sung and Collings, 2010] is considered; based on a reformulation of (6.56). This approach guarantees an improvement in the performance over the standard high-SNR approximation in fading channels [Sung and Collings, 2010]. In order to obtain the reformulation, the following bound obtained in [Papandriopoulos et al., 2008] is used

$$\begin{aligned} a\log z + b &\leq \log(1+z), \\ a=\frac{z_0}{1+z_0} \;\text{ and }\; b &= \log(1+z_0)-\frac{z_0}{1+z_0}\log z_0, \end{aligned} \tag{6.58}$$

for some $z_0\geq 0$, with equality when $z=z_0$.

Lemma 6.5. *With the change of variables $\widetilde{p}_k=\log p_k,\ k=1\ldots,K$, the lower bound*

$$\begin{aligned} \frac{a_k}{\log 2}\log\left(\frac{e^{\widetilde{p}_k}|\boldsymbol{h}_k^\dagger\boldsymbol{w}_k|^2}{\sum_{j\neq k}e^{\widetilde{p}_j}|\boldsymbol{h}_k^\dagger\boldsymbol{w}_j|^2+\sigma^2}\right)+\frac{b_k}{\log 2} \\ \leq \log_2\left(1+\frac{p_k|\boldsymbol{h}_k^\dagger\boldsymbol{w}_k|^2}{\sum_{j\neq k}p_j|\boldsymbol{h}_k^\dagger\boldsymbol{w}_j|^2+\sigma^2}\right), \end{aligned} \tag{6.59}$$

is concave in $\widetilde{p}_k,\ k=1,\ldots,K$.

Proof: The result follows immediately using the method in Lemma 6.4. ■

In Lemma 6.4 it was shown that the second term of (6.54) is concave by the second order condition. By using the lower bound in (6.59) for the first term of (6.54), a concave objective function is obtained. Since the constraints are affine, the optimization problem

arising from (6.56) and the bound (6.59) is a convex optimization problem. This convex optimization problem is given by

$$\begin{aligned} \underset{\widetilde{\mathbf{p}}}{\text{maximize}} \quad & \sum_{k=1}^{K}\left[\frac{a_k}{\log 2}\log\left(\frac{e^{\widetilde{p}_k}|\mathbf{h}_k^{\dagger}\mathbf{w}_k|^2}{\sum_{j\neq k}e^{\widetilde{p}_j}|\mathbf{h}_k^{\dagger}\mathbf{w}_j|^2+\sigma^2}\right)\right. \\ & \left.+\frac{b_k}{\log 2}-\log_2\left(1+\frac{e^{\widetilde{p}_k}\sum_{j\neq k}|\mathbf{h}_j^{\dagger}\mathbf{w}_k|^2}{\sigma^2}\right)\right] \\ \text{subject to} \quad & \text{tr}\left\{\mathbf{W}_{\text{p}}^{\dagger}\mathbf{W}_{\text{p}}\right\}\leq 1 \end{aligned} \tag{6.60}$$

The power allocation vector can then be obtained using Algorithm 1 in Table 6.1. To show that Algorithm 1 converges monotonically to a local optimum, note the constraint is the same for both the t-th and $(t+1)$-th subproblems. Hence, the solution of the t-th subproblem (6.60) is also feasible for the $(t+1)$-th subproblem (6.60). Moreover, by the bound in (6.58), the objective function is monotonically increasing and converges to a local optimum.

6.5.3 Proposed Precoding Scheme

Having established an algorithm to determine the optimal power allocation vector $\mathbf{p}$ for a fixed α, the precoding scheme is obtained by considering the joint optimization of α and $\mathbf{p}$. The joint optimization problem can be written as

$$\begin{aligned} \underset{\mathbf{p},\alpha}{\text{maximize}} \quad & R_{\text{s}}^{\text{PA}}(\mathbf{p},\alpha) \\ \text{subject to} \quad & \text{tr}\left\{\mathbf{W}_{\text{p}}^{\dagger}\mathbf{W}_{\text{p}}\right\}\leq 1. \end{aligned} \tag{6.61}$$

Even after using the transformation $\widetilde{p}_k=\log p_k,\ k=1,\ldots,K$, the problem (6.61) is non-convex. To solve this problem, Algorithm 2 is proposed in Table 6.1.

At each iteration, Algorithm 2 optimizes the regularization parameter α and subsequently the power allocation vector $\mathbf{p}$. It is straightforward to prove that Algorithm 2 converges monotonically and it thus provides with a locally optimal pair $(\alpha,\mathbf{p})$ for the proposed linear precoder. In the next section, simulations will show that the proposed precoder with jointly optimal regularization parameter and power allocation vector out-

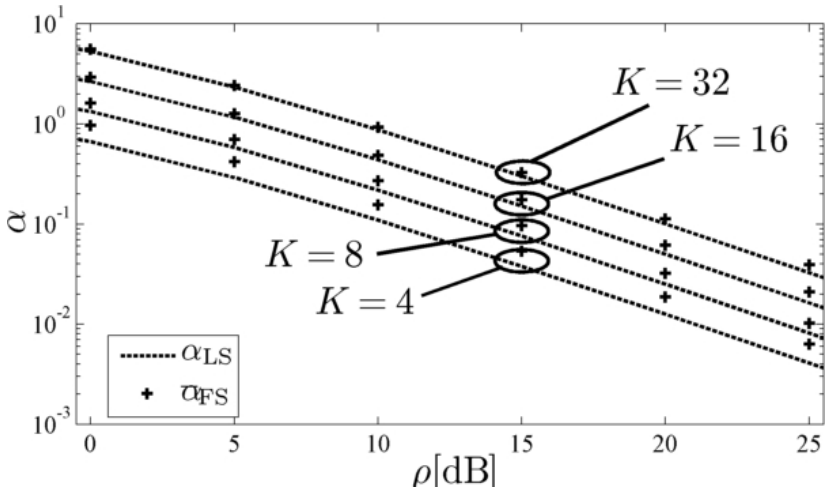


Figure 6.2: Comparison between the large-system regularization parameter α_{LS} and the value $\overline{\alpha}_{\mathrm{FS}}$ that maximizes the average secrecy sum-rate for finite K.

performs RCI precoding with α_{LS} and equal power allocation (RCI-EP).

6.6 Numerical Results

In this section, the performance of the proposed precoding scheme is described via simulations. The finite user scenario is also considered to show that many results from the large-system analysis hold for a small number of users. The precoding matrix $\mathbf{W}$ was normalized by $\sqrt{\gamma}$, as in (6.23), in order to meet the power constraint in (6.4). This corresponds to a long-term power constraint, which does not require the receivers to know the instantaneous value of γ [Peel et al., 2005]. In the following, $\alpha_{\mathrm{LS}} = K\xi_{\mathrm{opt}}$ denotes the large-system regularization parameter, obtained from (6.34).

Fig. 6.2 compares the large-system regularization parameter α_{LS} to the optimal regularization parameter $\overline{\alpha}_{\mathrm{FS}}$ for a finite number of users. The value of $\overline{\alpha}_{\mathrm{FS}}$ was found by using single-variable numerical optimization to maximize the mean value of the secrecy sum-rate in (6.28). The figure shows the finite-system and large-system regularization parameters at practical SNR values for four different numbers of users: 4, 8, 16 and 32. Observe that as the number of users K increases, the value of $\overline{\alpha}_{\mathrm{FS}}$ approaches the large-system regularization parameter α_{LS}.

In Fig. 6.3 it is demonstrated that using the large-system regularization parameter α_{LS}

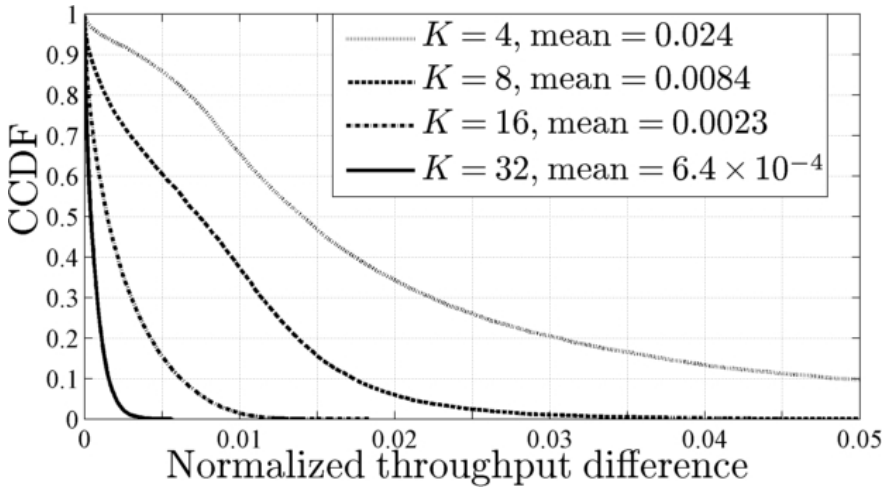


Figure 6.3: Complementary cumulative distribution function (CCDF) of the normalized secrecy sum-rate difference between using $\alpha_{\mathrm{FS}}(\mathbf{H})$ and α_{LS} with $\rho = 10$dB.

in a finite-size system does not cause a significant loss in the secrecy sum-rate compared to using a regularization parameter $\alpha_{\mathrm{FS}}(\mathbf{H})$ optimized for each channel realization. The figure shows the complementary cumulative distribution function (CCDF) of the normalized secrecy sum-rate difference between using α_{LS} and $\alpha_{\mathrm{FS}}(\mathbf{H})$ as the regularization parameter of the RCI precoder for $K = 4$, 8, 16, 32 users at an SNR of 10dB. The difference is normalized by dividing by the secrecy sum-rate of the precoder that uses $\alpha_{\mathrm{FS}}(\mathbf{H})$. Observe that the average normalized secrecy sum-rate difference is less than 2.4 percent for all values of K. As a result, the large-system regularization parameter α_{LS} may be used instead of the finite-system regularization parameter with only a small loss of performance. Moreover, the value of α_{LS} does not need to be calculated for each channel realization.

Fig. 6.4 compares the analytical secrecy sum-rate of the RCI precoder in (6.38) to the simulated secrecy sum-rate of the RCI precoder with a finite number of users, which is averaged over 10^3 channels. The RCI precoder with a finite number of users was obtained by using the regularization parameter $\overline{\alpha}_{\mathrm{FS}}$, found by simulation, that maximizes the average secrecy sum-rate. Observe that the large-system analysis is accurate at low SNR for all values of K. Moreover as K increases, the large-system analysis is accurate for larger values of the SNR.

In Fig. 6.5 the simulated secrecy sum-rate of the RCI precoder using the large-system

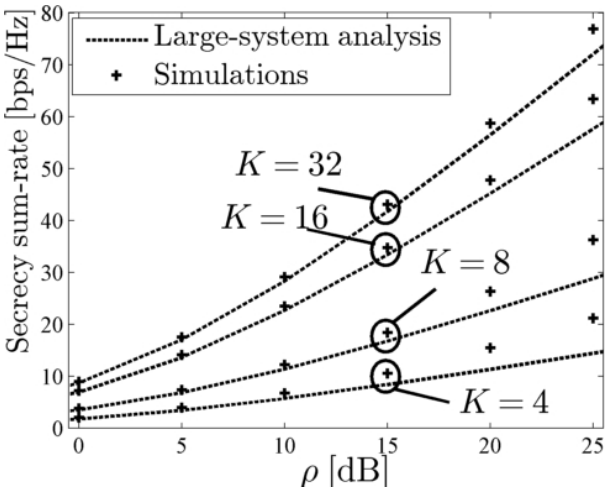
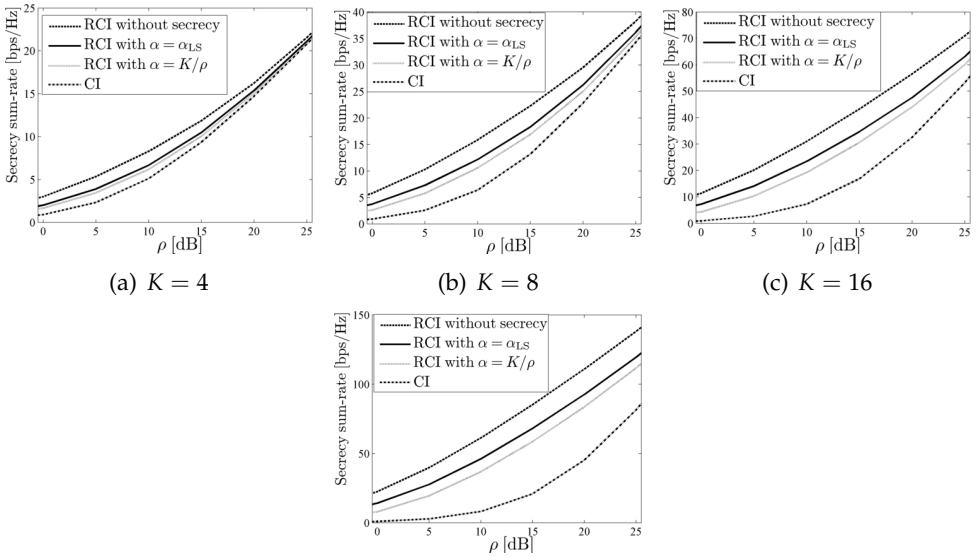


Figure 6.4: Comparison between the secrecy sum-rate in the large-system regime (6.38) and the simulated secrecy sum-rate for finite K.

regularization parameter α_{LS} is compared with CI precoding [Yoo and Goldsmith, 2006] and RCI precoding with $\alpha = K/\rho$, which maximizes the sum-rate without secrecy [Peel et al., 2005]. The sum-rate of the optimal RCI precoder without secrecy requirements is also plotted. The figure shows plots for $K = 4$, 8, 16, 32. Observe that CI precoding exhibits a large performance loss compared to the secrecy sum-rate of the optimal RCI precoder for large values of K. The RCI precoder with $\alpha = K/\rho$ outperforms CI precoding, but it is suboptimal compared to the RCI precoder that uses α_{LS}. Note that although CI precoding achieves secrecy in a simple way by completely canceling the information leakage, this comes at the cost of a poor sum-rate. Secrecy can be achieved with a significantly larger sum-rate by using the RCI precoder with α_{LS}. Also observe that the secrecy loss between the sum-rate of the RCI precoder without secrecy and the secrecy sum-rate of the RCI precoder is almost constant at high SNR for large K. This confirms the result derived in (6.50). Moreover, the value of the simulated per-antenna secrecy loss is 0.59 bits for $K = 32$ and $\rho = 25$dB; close to the 0.62 bits suggested by the analysis in (6.50).

In Fig. 6.6 the simulated per-user secrecy rate of the proposed precoder is compared with jointly optimized regularization parameter α and power allocation vector $\mathbf{p}$ to the RCI precoder with α_{LS} and $\mathbf{p}_{\text{opt}}$, and to the RCI-EP precoder with α_{LS}. Observe that there is a negligible performance difference between the proposed precoder and the RCI precoder with α_{LS} and $\mathbf{p}_{\text{opt}}$. As a result, a low-complexity, near-optimal RCI precoder may



(a) $K = 4$ (b) $K = 8$ (c) $K = 16$

(d) $K = 32$

Figure 6.5: Comparison between the RCI precoder with α_{LS} and other linear schemes. The secrecy loss is also shown as the gap between dashed and solid lines.

be implemented by using α_{LS} and optimizing the power vector separately. The figure shows that for $K = 4$, the proposed power allocation scheme always outperforms the RCI-EP precoder with α_{LS} by up to 20 percent, and the gain does not vanish at high SNR. This occurs because at high SNR $\xi_{\mathrm{opt}} \to 0$ and the RCI precoder behaves as a CI precoder, for which the optimal power allocation is waterfilling [Yoo and Goldsmith, 2006]. Hence, equal power allocation for RCI is suboptimal at high SNR. Fig. 6.6 also shows that the proposed power allocation scheme reduces the sum-rate loss due to the secrecy requirements. For $\rho \geq 15$dB, RCI with power allocation achieves a per-user secrecy rate which is even higher than the per-user rate achieved by the optimal RCI-EP without secrecy requirements. Furthermore, Fig. 6.6 shows the simulated secrecy capacity C_s^{MISOME} of a MISOME channel with the same per-message transmitted power. Although C_s^{MISOME} is obtained in a single-user and interference-free system [Khisti and Wornell, 2010a], at high SNR, RCI with power allocation achieves a per-user secrecy rate as large as C_s^{MISOME}.

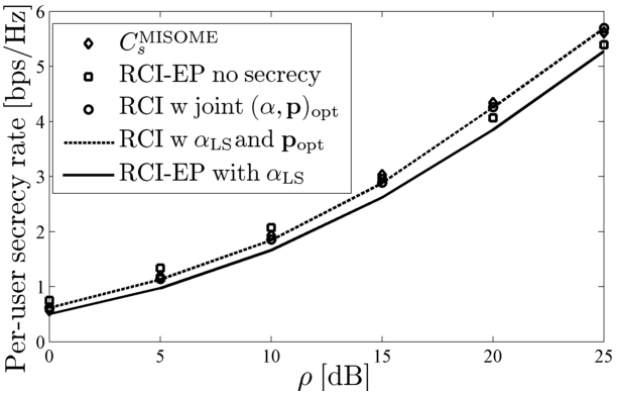


Figure 6.6: Per-user secrecy rate vs. ρ for $K = 4$ users: without power allocation (solid), with α_{LS} and $\mathbf{p}_{\text{opt}}$ (dashed) and with joint optimal $(\alpha, \mathbf{p})_{\text{opt}}$ (circle). The rate of the optimal RCI precoder without secrecy requirements (square) and the secrecy capacity of the MISOME channel (diamond) are also plotted.

6.7 Conclusions

In this chapter, the problem of secret communication in a MU-MIMO system with malicious users was considered. A linear precoder based on regularized channel inversion (RCI) was proposed with a regularization parameter and power allocation vector that maximize the achievable secrecy sum-rate. The analysis presented in the chapter, as well as the simulation results, showed that RCI with equal power allocation (RCI-EP) and with the optimal regularization parameter outperforms several other linear precoding schemes. Moreover, it achieves a sum-rate that has same scaling factor as the sum-rate of the optimum RCI precoder without secrecy requirements. The secrecy requirements result in a loss in terms of the sum-rate. This loss can be compensated by the proposed power allocation scheme, which increases the secrecy sum-rate compared to RCI-EP.

Chapter 7
Conclusions

7.1 Summary of Contributions

THIS thesis focused on areas that are expected to play an important role in future wireless multiuser and heterogeneous networks. In particular, three specific design challenges were addressed:

1. cross-layer scheduling;
2. structured limited feedback codebook design;
3. and resource allocation for physical layer security.

The first design challenge addressed in this thesis was cross-layer scheduling in multiuser networks. Two network architectures were considered: single-cell networks, aided by a relay (Chapter 2); and the CoMP network (Chapter 3). In both networks, heuristic cross-layer schedulers were proposed based on the maximum weighted SNR policy, with the weight chosen to be a function of the delay in packet scheduling. Scheduling policies with the same structure were able to be applied to both the user scheduler in Chapter 2 and the BS scheduler in Chapter 3 due to the centralized structure of the scheduling policy. Importantly, the trends observed for the user scheduler in the single-cell also apply for the BS scheduler in the CoMP network, which is a consequence of the common structure for both scheduling policies.

Three key results were obtained for the cross-layer schedulers in Chapter 2 and 3. First, it was shown that there is up to a 30% improvement in the delay was achieved using the cross-layer policy compared to the opportunistic (or equal weight) policy, while

achieving a comparable SEP for the scheduled user. This is surprising as the SEP of the opportunistic policy forms a lower bound, and hence the cross-layer policy can achieve a near optimal SEP.

Second, the optimal code rate in coded systems is strongly dependent on the transmission time of the uncoded system. In particular, there are three regimes: SEP dominated, where the throughput is dominated by the effect of the SEP and a low code rate is optimal; PLP dominated, where the throughput is dominated by the PLP and a high code-rate is optimal; and the intermediate regime, where the throughput is dominated by neither the PLP nor the SEP, and the optimal code-rate is approximately $1/2$.

Third, the achievable PLP for the fixed weight scheduling policy is comparable with the adaptive weight policy, under a wide range of conditions. In particular, this occurs when there is a high probability of scheduling a given user over the others, and for moderate and long transmission times. This result is important in CoMP as CSI and delay side information is shared over finite-bandwidth backhaul links to the centralized scheduler. The required backhaul bandwidth is then reduced when the instantaneous packet delay is not required at the scheduler, as is the case for the fixed weight policy.

The second design challenge addressed in this thesis was structured limited feedback codebook design. Two particular problems were addressed: the construction of different codebooks at each user to avoid a rank-deficient quantized channel matrix and hence a low sum-rate with the ZF precoder; and low-complexity CoMP codebook constructions to adapt to changes in the number of cooperating BSs. Although there are important differences between the two problems, there are two key unifying concepts.

The first unifying concept is that the Fourier codebook is an important practical codebook construction. In Chapter 4, it was shown that using the proposed Householder construction, the resulting transformation of the Fourier codebook is up to 50% sparse. This means that the search complexity can be reduced by up to half. In Chapter 5, the Fourier codebook was used as the basis of a low-complexity construction to adapt to varying numbers of cooperating BSs in CoMP. In particular, the proposed construction was shown to have linear complexity in both size and dimension, compared with the standard approach, which has exponential complexity.

The second unifying concept is the applicability of theory of equiangular tight frames in codebook construction. It has been long known that equiangular tight frames are optimal with respect to the important Grassmannian design criterion. In Chapter 4, it was shown that all real, and group covariant equiangular tight frames also optimize the rate-outage lower bound in multiuser MIMO. Equiangular tight frames were then exploited in Chapter 5 to construct variable dimension codebooks with low storage, which was achieved by using difference set decompositions.

Chapter 4 also provided a link between the role of efficient codebook constructions and user scheduling. In particular, it was shown that the proposed construction of different codebooks at each user improved the probability that the target rates for each user were met by up to 15% compared with user scheduling. Surprisingly, this means that more sophisticated quantization codebook construction can improve performance more than introducing additional scheduling within the active user group.

The third design challenge addressed in this thesis was resource allocation to achieve physical layer security in MU-MIMO. In particular, the design of a RCI-based precoder and power allocation were considered in Chapter 6. Two key results were obtained. First, it was shown that providing secrecy resulted in a 3.75 dB power loss, for high SNR in the large-system regime under equal power allocation. This result shows that efficient resource allocation is an important design problem in systems providing secrecy as the required power is more than double that of standard systems to achieve the same rate. Second, it was shown that the regularization parameter for the RCI precoder and the power control can be optimized separately. This is due to the fact that the secrecy sum-rate achieved by the joint precoder and power control design in Chapter 6 was comparable with separate optimization. As such, the complexity of the resource allocation algorithm can be considerably reduced.

7.2 Future Research

The work in this thesis motivates a number of future research directions. The first new direction is a more extensive application of the proposed concept of τ-achievable packet

loss probability region to different networks. At this point, it is unclear if instantaneous packet delay information is necessary to achieve high performance, as exhibited in Chapter 3. If this is the case, the amount of side information required to be shared in large scale CoMP and heterogeneous network can be significantly reduced by exploiting long-term statistics of the arrival and service processes.

While the effect of delay and packet loss probability was investigated in Chapter 2 and 3 for single cell and CoMP networks, the effect of these metrics in general heterogeneous networks remains an open problem. As video data is an important part of heterogeneous network traffic and these metrics largely determine the performance, this is an important avenue to investigate. Some initial work in this direction is the paper [Egan and Collings, 2013], where different frequency reuse schemes were compared under minimum rate constraints and queue stability conditions.

There are several interesting open problems related to the structured codebook design problems investigated in Chapters 4 and 5. Notably, equally likely codewords appear to play an important role in the construction of high resolution codebooks. This is particularly evident in the construction of optimal base codebooks based on the rate-outage lower bound in Chapter 4. Codebooks with equally likely codewords also appeared as one of the minimizers of the expected square correlation described in Chapter 5, and also detailed in [Egan et al., 2011]. This suggests an intimate link between the action of unitary representations of groups and high performance codebook design.

In Chapter 5, decompositions of difference sets were exploited to construct low storage variable dimension codebooks for CoMP. At present, a full classification of these decompositions is not available so there is work remaining in this direction. These decompositions also play an important role in the construction of variable rate CDMA and optical codes. As such the construction of difference set decompositions is a well motivated problem.

Physical layer security has developed into a large and important area of research, with many generalizations of the early work of Wyner. From the perspective of this thesis, new work has emerged extending the results in Chapter 6 to networks with the number of users different to the number of transmit antennas [Geraci et al., a]. This has since been

further extended to heterogeneous cellular networks with BSs distributed according to a Poisson process [Geraci et al., b]. Although many analytical properties of these networks have by now been well investigated, the resource allocation aspect of the problem in terms of power control and scheduling remains an important avenue of research. In particular, since the secrecy required up to twice the transmit power to achieve the same performance as standard multiuser MIMO (see Chapter 6) additional resource allocation and signal processing should be considered to improve network efficiency.

Appendix A
Proofs for Chapter 2

A.1 Proof of Lemma 2.1

Without loss of generality, consider the case where $\overline{\gamma}_{Rm} = 1,\ m = 1, 2, \ldots, K$. The probability that the k-th user is selected in state $\mathbf{s}$ (corresponding to weights $W_k = W_k(s_k)$) is given by

$$P_k(\mathbf{s}) = \int_0^\infty \Pr(W_k\gamma_{Rk} > x | \Gamma_{k^-} = x) f_{\Gamma_{k^-}}(x) dx, \tag{A.1}$$

where $f_{\Gamma_{k^-}|\mathbf{s}}(x|\mathbf{s})$ is the pdf of $\Gamma_{k^-} = \max\{W_m\gamma_{Rm}\}_{m=1,m\neq k}^K$. The pdf of Γ_{k^-} is derived by differentiating

$$\Pr(\Gamma_{k^-} \leq x|\mathbf{s}) = \prod_{\substack{m=1\\m\neq k}}^K \Pr(W_m\gamma_{Rm} \leq x) = \prod_{\substack{m=1\\m\neq k}}^K (1 - e^{-x/W_m}) \tag{A.2}$$

that results in

$$f_{\Gamma_{k^-}|\mathbf{s}}(x|\mathbf{s}) = \sum_{\substack{i=1\\i\neq k}}^K \frac{1}{W_i} e^{-x/W_i} \prod_{\substack{m=1\\m\neq i,k}}^K (1 - e^{-x/W_m}), \tag{A.3}$$

Substituting (A.3) into (A.1), noting that γ_{Rk} is exponentially distributed with $\overline{\gamma}_{Rk} = 1$, yields

$$P_k(\mathbf{s}) = \int_0^\infty e^{-x/W_k} \sum_{i=1,i\neq k}^K \frac{1}{W_i} e^{-x/W_i} \prod_{\substack{m=1\\m\neq i,k}}^K (1 - e^{-x/W_m})$$

$$\times \sum_{\substack{i=1 \\ i\neq k}}^{K} \frac{1}{W_i} \int_0^{\infty} e^{-x/W_k} e^{-x/W_i} \prod_{\substack{m=1 \\ m\neq i,k}}^{K} (1 - e^{-x/W_m}). \tag{A.4}$$

The product term in (A.4) may be expanded as

$$\prod_{\substack{m=1 \\ m\neq i,k}}^{K} (1 - e^{-x/W_m}) = 1 + \sum_{m=1}^{K-2} \sum_{\substack{n_p = n_{p-1}+1 \\ p=1,\ldots,m}}^{K-2-m+p} (-1)^m e^{\left(-\sum_{j=1}^{m} x\tilde{W}_{n_j}^{-1}\right)}, \tag{A.5}$$

where the weights are re-expressed as $(\tilde{W}_q)_{q=1}^{K-2} = (W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_{k-1}, W_{k+1}, \ldots, W_K)$. Substituting (A.5) in (A.4) and evaluating the integrals, the desired result is obtained.

A.2 Proof of Lemma 2.2

The conditional CDF of the second hop SNR is now derived. First, condition on the state vector $\mathbf{s}$ that results in weights $W_k = W_k(s_k)$. The CDF of the second hop channel gain $|h_{Rk}^*|^2$ of the scheduled user is evaluated according to

$$\begin{aligned} F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) &= \sum_{k=1}^{K} \Pr(|h_{Rk}|^2 \leq x, W_k|h_{Rk}|^2 > \Gamma_{k^-}) \\ &= \sum_{k=1}^{K} \int_0^{W_k x} \Pr(\frac{y}{W_k} < |h_{Rk}|^2 \leq x | \Gamma_{k^-} = y) \\ &\quad \times f_{\Gamma_{k^-}}(y) dy, \end{aligned} \tag{A.6}$$

where $f_{\Gamma_{k^-}}(y)$ is given in (A.3). Note that the formulation in terms of $|h_k|^2$ is a special case of γ_{Rk} where $\overline{\gamma}_{Rm} = 1,\ m = 1, 2, \ldots, K$. Applying (A.3) and (2.6), to obtain

$$\begin{aligned} F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) = \sum_{k=1}^{K} \sum_{\substack{i=1 \\ i\neq k}}^{K} \frac{1}{W_i} \int_0^{W_k x} \left[(1 - e^{-x}) - (1 - e^{-y/W_k})\right] \\ \times e^{-y/W_i} \prod_{\substack{m=1 \\ m\neq i,k}}^{K} (1 - e^{-y/W_m}) dy. \end{aligned} \tag{A.7}$$

$$F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) = \sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\frac{1}{W_i}\left[-e^{-x}W_i\left(1-e^{-W_k x/W_i}\right)+\frac{1}{\frac{1}{W_i}+\frac{1}{W_k}}\left(1-e^{-W_k x(1/W_i+1/W_k)}\right)\right]$$
$$+\sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p}\frac{(-1)^m}{W_i}\left[-\frac{e^{-x}\left(1-e^{-W_k x\left(\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\right)}{\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}}\right.$$
$$\left.+\frac{\left(1-e^{-W_k x\left(\frac{1}{W_k}+\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\right)}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_k}+\frac{1}{W_i}}\right] \tag{A.9}$$

Using the same product expansion as in (A.5) results in

$$F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) = \sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\frac{1}{W_i}\int_0^{W_k x}dy\left[(1-e^{-x})-(1-e^{-y/W_k})\right]$$
$$\times e^{-y/W_i}\left(1+\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p}(-1)^m e^{-\sum_{j=1}^{m}y\tilde{W}_{n_j}^{-1}}\right). \tag{A.8}$$

After a number of simple integrations and algebraic simplifications, (A.9) is obtained.

To obtain the CDF of the second hop SNR γ_2, make the substitution $x = \frac{\gamma}{\bar{\gamma}_2}$.

A.3 Proof of Lemma 2.3

Substituting the conditional CDF of γ_2 in Lemma 2.2 and the PDF of γ_1 in (2.6) into (2.19), yields the following expression for the conditional CDF of the end-to-end SNR, given by

$$F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s}) = 1-e^{-\frac{\gamma}{\bar{\gamma}_1}}+\sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\frac{1}{W_i}\left(-W_i\Phi_1+W_i\Phi_2+\Phi_3-\Phi_4\right)$$
$$+\sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p}\frac{(-1)^m}{W_i}\left(-\Phi_5+\Phi_6+\Phi_7-\Phi_8\right), \tag{A.10}$$

where

$$\Phi_1 = \frac{1}{\overline{\gamma}_1} e^{-\frac{\gamma}{\overline{\gamma}_1}} \int_0^\infty e^{-\frac{\omega}{\overline{\gamma}_1}} e^{-\frac{1}{\overline{\gamma}_2}\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)} d\omega,$$

$$\Phi_2 = \frac{1}{\overline{\gamma}_1} e^{-\frac{\gamma}{\overline{\gamma}_1}} \int_0^\infty e^{-\frac{\omega}{\overline{\gamma}_1}} e^{-\left(\frac{1}{\overline{\gamma}_2}+\frac{W_k}{\overline{\gamma}_2 W_i}\right)\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)} d\omega,$$

$$\Phi_3 = \frac{1}{\overline{\gamma}_1} e^{-\frac{\gamma}{\overline{\gamma}_1}} \int_0^\infty e^{-\frac{\omega}{\overline{\gamma}_1}} \frac{W_i W_k}{W_i + W_k} d\omega,$$

$$\Phi_4 = \frac{e^{-\frac{\gamma}{\overline{\gamma}_1}}}{\overline{\gamma}_1} \int_0^\infty \frac{e^{-\frac{\omega}{\overline{\gamma}_1}} W_i W_k}{W_i + W_k} e^{-\frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i}+\frac{1}{W_k}\right)\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)} d\omega,$$

$$\Phi_5 = \frac{1}{\overline{\gamma}_1 \left(W_i^{-1} + \sum_{j=1}^m \tilde{W}_{n_j}^{-1}\right)},$$

$$\Phi_6 = \frac{e^{-\frac{\gamma}{\overline{\gamma}_1}}}{\overline{\gamma}_1} \int_0^\infty \frac{e^{-\frac{\omega}{\overline{\gamma}_1}} e^{-\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)\left(\frac{1}{\overline{\gamma}_2}+\frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i}+\sum_{j=1}^m \tilde{W}_{n_j}^{-1}\right)\right)}}{\frac{1}{W_i} + \sum_{j=1}^m \tilde{W}_{n_j}^{-1}} d\omega,$$

$$\Phi_7 = \frac{e^{-\frac{\gamma}{\overline{\gamma}_1}}}{\sum_{j=1}^m \tilde{W}_{n_j}^{-1} + \frac{1}{W_k} + \frac{1}{W_i}},$$

$$\Phi_8 = \frac{1}{\overline{\gamma}_1} \int_0^\infty \frac{e^{-\frac{\omega}{\overline{\gamma}_1}} e^{-\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)\frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_k}+\frac{1}{W_i}+\sum_{j=1}^m \tilde{W}_{n_j}^{-1}\right)}}{\sum_{j=1}^m \tilde{W}_{n_j}^{-1} + \frac{1}{W_i} + \frac{1}{W_k}} d\omega.$$

The integrals in Φ_1, Φ_2, Φ_3, Φ_4, Φ_6, and Φ_8 are solved using the identity

$$\int_0^\infty e^{-\frac{\omega}{a}-\frac{b}{\omega}} d\omega = 2\sqrt{\frac{a}{b}} K_1\left(2\sqrt{\frac{b}{a}}\right),$$

where $K_1(x)$ is the modified Bessel function of the second kind.

A.4 Proof of Theorem 2.2

Observe that if there is a new packet arrival at time t, the PLP of a given user can be expressed as $P_L = \Pr(\text{buffer is full for new arrival at time } t)$. Applying Lemma 2.4 and conditioning on the arrival time t and the number of packets in the buffer l, the PLP is

approximated as (which is an exact result for fixed weights)

$$P_L \approx \sum_{l=0}^{B} \int_0^T \sum_{m=B-l}^{\infty} e^{-\lambda t} \frac{(\lambda t)^m}{m!} \nu_l \frac{1}{T} dt \tag{A.11}$$

The final result is obtained by using (2.1) and solving the integral using eq. 3.351 in [Gradshteyn and Ryzhik, 2007]

Appendix B

Proofs for Chapter 4

B.1 Proof of Theorem 4.1

In order to prove the theorem, it is necessary that there exists no codewords in $\{\phi_i\}_{i=1}^N \cup \{\mathbf{U}_H\phi_i\}_{i=1}^N$ that lie in the same one-dimensional subspace. First, there must be no codewords in $\{\phi_i\}_{i=1}^N$ that lie in the same subspace. This is achieved by requiring that $\phi_i \neq e^{j\theta}\phi_k$, which is the first condition in the theorem. Since $\mathbf{U}_H$ is an isometry, this condition also ensures that no codewords in $\{\mathbf{U}_H\phi_i\}_{i=1}^N$ lie in the same subspace.

Second, there must be no codewords such that $\mathbf{U}_H\phi_i = e^{j\theta}\phi_k$. Note that $\phi_i = \mathbf{U}_{ki}\phi_k$, so this is equivalent to $\mathbf{U}_H\mathbf{U}_{ki}\phi_k = e^{j\theta}\phi_k$. To ensure that there are no such codewords, it is required that ϕ_k is not an eigenvector of $\mathbf{U}_H\mathbf{U}_{ki}$. This is precisely the second condition of the theorem.

Since these are the only possible scenarios where codewords in $\{\phi_i\}_{i=1}^N \cup \{\mathbf{U}_H\phi_i\}_{i=1}^N$ can lie in the same subspace, the theorem is proved and the codebooks are unique.

B.2 Proof of Proposition 4.2

Consider a 2-dimensional frame with elements that are roots of unity. This means that all frame elements are of the form $\phi_i = \frac{1}{\sqrt{2}}[e^{j\frac{l_1\pi}{N}}, e^{j\frac{l_2\pi}{N}}]^T$. Suppose a Householder transform is chosen such that $\mathbf{H}\phi_j = \frac{1}{\sqrt{2}}[1,0]^T,\ j \neq i$, where $\phi_j = \frac{1}{\sqrt{2}}[e^{j\frac{k_1\pi}{N}}, e^{j\frac{k_2\pi}{N}}]^T$.

In this case, the following is obtained

$$\mathbf{H}\phi_i = \frac{1}{2}\begin{pmatrix} e^{-j\frac{k_1\pi}{N}}e^{j\frac{l_1\pi}{N}} + e^{-j\frac{k_2\pi}{N}}e^{j\frac{l_2\pi}{N}} \\ e^{-j\frac{k_1\pi}{N}}e^{j\frac{l_1\pi}{N}} - e^{-j\frac{k_2\pi}{N}}e^{j\frac{l_2\pi}{N}} \end{pmatrix} \tag{B.1}$$

In order to improve the sparsity of the transformed frame element $\mathbf{H}\phi_i$, a scalar transformation $e^{j\theta}$ is applied such that

$$e^{j\theta} = \frac{e^{j\frac{k_1\pi}{N}}e^{-j\frac{l_1\pi}{N}} + e^{j\frac{k_2\pi}{N}}e^{-j\frac{l_2\pi}{N}}}{|e^{j\frac{k_1\pi}{N}}e^{-j\frac{l_1\pi}{N}} + e^{j\frac{k_2\pi}{N}}e^{-j\frac{l_2\pi}{N}}|}. \tag{B.2}$$

Immediately, observe that the first element of $e^{j\theta}\mathbf{H}\phi_i$ is real.

As only the phase is of interest, not the magnitude of the second element after the transformations, consider the unnormalized second element

$$v = \left(e^{j\frac{k_1\pi}{N}}e^{-j\frac{l_1\pi}{N}} + e^{j\frac{k_2\pi}{N}}e^{-j\frac{l_2\pi}{N}}\right)\left(e^{-j\frac{k_1\pi}{N}}e^{j\frac{l_1\pi}{N}} - e^{-j\frac{k_2\pi}{N}}e^{j\frac{l_2\pi}{N}}\right). \tag{B.3}$$

Several algebraic manipulations yields

$$v = 2j\left(\sin\frac{\pi}{N}(l_1-k_1)\cos\frac{\pi}{N}(l_2-k_2) - \sin\frac{\pi}{N}(l_2-k_2)\cos\frac{\pi}{N}(l_1-k_1)\right), \tag{B.4}$$

which is imaginary. As a result, the transformed elements are either real or imaginary except for the first codeword given by $\frac{1}{\sqrt{2}}[1,0]^T$. The density is then given by

$$\kappa = \frac{\left(1+\frac{2(N-1)N_t}{2}\right)}{2N_tN} = \frac{2N-1}{4N}. \tag{B.5}$$

B.3 Proof of Proposition 4.3

First observe that

$$\Pr(\hat{\mathbf{H}} \text{ invertible}) = 1 - \Pr(\hat{\mathbf{H}} \text{ not invertible}). \tag{B.6}$$

Using a similar argument as for (4.21), it is straightforward to show that $\Pr(\hat{\mathbf{H}} \text{ not invertible})$ is a posynomial. The optimization problem

$$\begin{aligned} \min_{p_i,\, i=1,\ldots,N} \quad & \Pr(\hat{\mathbf{H}} \text{ not invertible}) \\ \text{subject to} \quad & 0 \leq p_i,\ \forall\, i = 1, 2, \ldots, N \\ & \sum_{i=1}^{N} p_i = 1, \end{aligned} \tag{B.7}$$

can then be transformed to a convex optimization problem using the standard method for geometric programs (it is minimizing instead of maximizing). Since the constraints are linear and a feasible solution exists, the KKT conditions are necessary and sufficient for global optimality, by Slater's condition [Boyd and Vandenberghe, 2004]. This means that all solutions of the KKT conditions give the same optimal value for the objective.

Now turn to (4.21). Using the KKT conditions, the derivatives of the Lagrangian are given by

$$\frac{\partial L}{\partial p_k} = \nu - \lambda_k + \sum_{j=1}^{\sigma} \prod_{\substack{i \in S_j \\ i \neq k}} \mathbf{1}_{k \in S_j} p_i = 0, \tag{B.8}$$

where ν and $\lambda_m,\ m = 1, \ldots, N$ are the Lagrange multipliers, the other terms are defined as for (4.21), and $\mathbf{1}$ is the indicator function. Also note that ν acts as a slack variable, which means that it is zero at the optimal solution. Solving the equations in (B.8) and using the constraint that the probabilities sum to one, yields $p_1 = \cdots = p_N = \frac{1}{N}$.

Since the constraints in (4.21) are linear, it follows that the solution is a local optimum. As (B.6) holds, the local optimum of (4.21) is equivalent to the local optimum of (B.7). The local optimum is then a global optimum as (B.7) is equivalent to a convex optimization problem.

B.4 Proof of Theorem 4.2

Note that if $\mathbf{F}$ is real, then the Voronoi regions were shown to be congruent in [Mondal et al., 2007]. As a result, the Voronoi regions have the same volume and each codeword

is equally likely.

More generally in the case when the ETF is group-covariant, denote Ω^d as the unit sphere in $\mathbb{C}^d$ and $\mu(V_i)$ as the measure on $\mathbb{C}^d$ induced by the Haar measure on Ω^d for the Voronoi region V_i of codeword i. For codeword $i \in \{1, 2, \ldots, N\}$, then

$$\begin{aligned}\mu(V_i) &= \mu(\{\mathbf{z} \in \Omega^d || \mathbf{z}^\dagger \mathbf{U}_i \mathbf{f}_1|^2 > |\mathbf{z}^\dagger \mathbf{U}_j \mathbf{f}_1|^2,\ j \neq i\}) \\ &= \mu(\{\mathbf{z} \in \Omega^d || \mathbf{z}^\dagger \mathbf{f}_1|^2 > |\mathbf{z}^\dagger \mathbf{U}_i^\dagger \mathbf{U}_j \mathbf{f}_1|^2\ j \neq i\}) \\ &= \mu(\{\mathbf{z} \in \Omega^d || \mathbf{z}^\dagger \mathbf{f}_1|^2 > |\mathbf{z}^\dagger e^{j\zeta} \mathbf{U}_k \mathbf{f}_1|^2\ k \neq 1\}) = \mu(V_1), \end{aligned} \tag{B.9}$$

where the second equality follows since the Haar measure is invariant to unitary transformations. This shows that the Voronoi regions of each codeword all have the same volume. Hence, each codeword is equally likely.

Bibliography

[Akyildiz et al., 2007] Akyildiz, I., Melodia, T., and Chowdhury, K. (2007). Wireless multimedia sensor networks: a survey. *IEEE Wireless Communications*, 14(6):32–39.

[Andrews, 2013] Andrews, J. (2013). Seven ways that HetNets are a cellular paradigm shift. *IEEE Communications Magazine*, 51(3).

[Andrews et al., 2013] Andrews, J., Singh, S., Ye, Q., Lin, X., and Dhillon, H. (2013). An Overview of Load Balancing in HetNets: Old Myths and Open Problems. *arXiv:1307.7779*.

[Andrews et al., 2012] Andrews, J. G., Claussen, H., Dohler, M., Rangan, S., and Reed, M. C. (2012). Femtocells: past, present, and future. *IEEE Journal on Selected Areas in Communications*, 30(3):497–508.

[Anton-Haro et al., 2006] Anton-Haro, C., Svedman, P., Bengtsson, M., Alexiou, A., and Gameiro, A. (2006). Cross-layer scheduling for multi-user MIMO systems. *IEEE Communications Magazine*, 44(9):39–45.

[Berry and Gallager, 2002] Berry, R. and Gallager, R. (2002). Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory*, 48(5):1135–1149.

[Bhagavatula and Heath, 2011] Bhagavatula, R. and Heath, R. (2011). Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems. *IEEE Transactions on Signal Processing*, 59(2).

[Bolcskei et al., 2006] Bolcskei, H., Nabar, R., Oyman, O., and Paulraj, A. (2006). Capacity scaling laws in MIMO relay networks. *IEEE Transactions on Wireless Communications*, 5(6).

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

[Caire and Shamai, 2003] Caire, G. and Shamai, S. (2003). On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Transactions on Information Theory*, 49(7).

[Carneiro et al., 2004] Carneiro, G., Ruela, J., and Ricardo, M. (2004). Cross-layer design in 4G wireless terminals. *IEEE Wireless Communications*, 11(2):7–13.

[Casazza et al., 2011] Casazza, P., Heinecke, A., Krahmer, F., and Kutyniok, G. (2011). Optimally Sparse Frames. *IEEE Transactions on Information Theory*, 57(11).

[Chae et al., 2008] Chae, C.-B., Mazzarese, D., Jindal, N., and Heath, R. (2008). Coordinated beamforming with limited feedback in the MIMO broadcast channel. *Selected Areas in Communications, IEEE Journal on*, 26(8).

[Chandrasekhar et al., 2008] Chandrasekhar, V., Andrews, J., and Gatherer, A. (2008). Femtocell networks: a survey. *IEEE Communications Magazine*, 46(9):59–67.

[Chen and Hu, 2012] Chen, X. and Hu, R. (2012). Joint uplink and downlink optimal mobile association in a wireless heterogeneous network. In *Global Communications Conference*.

[Cheng et al., 2010] Cheng, Y., Lau, V., and Long, Y. (2010). A scalable limited feedback design for network MIMO using per-cell product codebook. *IEEE Transactions on Wireless Communications*, 9(10).

[Chiang et al., 2007a] Chiang, M., Low, S., Calderbank, A., and Doyle, J. (2007a). Layering as optimization decomposition: a mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1).

[Chiang et al., 2007b] Chiang, M., Tan, C., Palomar, D., Oneill, D., and Julian, D. (2007b). Power control by geometric programming. *IEEE Transactions on Wireless Communications*, 6(7):2640–2651.

[Choi et al., 2011] Choi, D., Lee, D., and Lee, J. (2011). Resource allocation for CoMP with multiuser MIMO-OFDMA. *IEEE Transactions on Vehicular Technology*, 60(9).

[Colbourn and Dinitz, 2007] Colbourn, C. and Dinitz, J. (2007). *The CRC Handbook of Combinatorial Designs*. CRC, 2nd edition.

[Conte et al., 2010] Conte, E., Tomasin, S., and Benvenuto, N. (2010). A comparison of scheduling strategies for MIMO broadcast channel with limited feedback on OFDM systems. *EURASIP Journal on Wireless Communications and Networking*, pages 1–12.

[Cover and El Gamal, 1979] Cover, T. and El Gamal, H. (1979). Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, 25(5).

[Csiszár and Körner, 1978] Csiszár, I. and Körner, J. (1978). Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348.

[Daley and Vere-Jones, 2003] Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume 1: Elementary Theory and Methods*, volume I. Springer, 2nd edition.

[Damnjanovic et al., 2011] Damnjanovic, A., Montojo, J., Wei, Y., Ji, T., Luo, T., Vajapeyam, M., Yoo, T., Song, O., and Malladi, D. (2011). A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications*, 18(3):10–21.

[Diggavi and Grossglauser, 2006] Diggavi, S. and Grossglauser, M. (2006). On information transmission over a finite buffer channel. *IEEE Transactions on Information Theory*, 52(3).

[Ding et al., 2011] Ding, H., Ge, J., da Costa, D. B., and Jiang, Z. (2011). A new efficient low-complexity scheme for multi-source multi-relay cooperative networks. *IEEE Transactions on Vehicular Technology*, 60(2):716–722.

[Ding et al., 2007] Ding, P., Love, D., and Zoltowski, M. D. (2007). Multiple antenna broadcast channels with shape feedback and limited feedback. *IEEE Transactions on Signal Processing*, 55(7).

[Egan and Collings, 2013] Egan, M. and Collings, I. (2013). Base station cooperation for queue stability in wireless heterogeneous cellular networks. In *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*.

[Egan et al., 2011] Egan, M., Sung, C., and Collings, I. (2011). Codebook Design for the Finite Rate MIMO Broadcast Channel with Zero-Forcing Precoding. In *IEEE GLOBECOM*.

[Egan et al., 2013a] Egan, M., Sung, C., and Collings, I. (2013a). Structured and Sparse Limited Feedback Codebooks for Multiuser MIMO. *IEEE Transactions on Wireless Communications*, 12(8).

[Egan et al., 2013b] Egan, M., Yeoh, P., Elkashlan, M., and Collings, I. (2013b). A new cross-layer user scheduler for wireless multimedia relay networks. *IEEE Transactions on Wireless Communications*, 12(1).

[Ephremides and Hajek, 1998] Ephremides, A. and Hajek, B. (1998). Information theory and communication networks: an unconsummated union. *IEEE Transactions on Information Theory*, 44(6).

[Eryilmaz and Srikant, 2004] Eryilmaz, A. and Srikant, R. (2004). Scheduling with QoS constraints over Rayleigh fading channels. In *2004 43rd IEEE Conference on Decision and Control (CDC)*, pages 3447–3452. IEEE.

[Fattah and Leung, 2002] Fattah, H. and Leung, C. (2002). An overview of scheduling algorithms in wireless multimedia networks. *IEEE Wireless Communications*, 9(5):76–83.

[Foschini and Gans, 1998] Foschini, G. and Gans, M. (1998). On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3).

[Gastpar et al., 2002] Gastpar, M., Kramer, G., and Gupta, P. (2002). The multiple-relay channel: coding and antenna clustering capacity. In *Proc. IEEE International Symposium on Information Theory*.

[Geraci et al., a] Geraci, G., Couillet, R., Yuan, J., Debbah, M., and Collings, I. Large system analysis of linear precoding in MISO broadcast channels with confidential messages. *to appear in IEEE Journal on Selected Areas in Communications*.

[Geraci et al., b] Geraci, G., Dhillon, H., Andrews, J., Yuan, J., and Collings, B. Physical layer security in downlink multi-antenna cellular networks. *submitted to IEEE Transactions on Communications*.

[Geraci et al., 2011] Geraci, G., Yuan, J., Razi, A., and Collings, I. B. (2011). Secrecy sum-rates for multi-user MIMO linear precoding. In *Proc. IEEE Int. Symp. on Wireless Commun. Systems (ISWCS)*.

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

[Gesbert et al., 2007] Gesbert, D., Kiani, S., Gjendemsjo, A., and Ien, G. (2007). Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks. *Proceedings of the IEEE*, 95(12).

[Ghosh et al., 2012] Ghosh, A., Mangalvedhe, N., Ratasuk, R., Mondal, B., Cudak, M., Visotsky, E., Thomas, T., Andrews, J., Xia, P., Jo, H., Dhillon, H., and Novlan, T. (2012). Heterogeneous cellular networks: From theory to practice. *IEEE Communications Magazine*, 50(6):54–64.

[Goel and Negi, 2008] Goel, S. and Negi, R. (2008). Guaranteeing secrecy using artificial noise. *IEEE Transactions on Wireless Communications*, 7(6).

[Gradshteyn and Ryzhik, 2007] Gradshteyn, I. and Ryzhik, I. (2007). *Table of Integrals, Series, and Products*. 7th edition.

[Han et al., 2009] Han, K., Choi, Y., Kim, D., Na, M., Choi, S., and Han, K. (2009). Optimization of femtocell network configuration under interference constraints. In *7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, Wireless Networks*.

[Han et al., 2007] Han, Z., Ji, Z., and Liu, K. (2007). Non-cooperative resource competition game by virtual referee in multi-cell OFDMA networks. *IEEE Journal on Selected Areas in Communications*, 25(6).

[Heath et al., 2009] Heath, R., Wu, T., and Soong, A. (2009). Progressive refinement of beamforming vectors for high-resolution limited feedback. *EURASIP Journal on Advances in Signal Processing*, page 13.

[Hoadley and Maveddat, 2012] Hoadley, J. and Maveddat, P. (2012). Enabling small cell deployment with HetNet. *IEEE Wireless Communications*, 19(2).

[Hochwald et al., 2000] Hochwald, B., Marzetta, T., Richardson, T., Sweldens, W., and Urbanke, R. (2000). Systematic design of unitary space-time constellations. *IEEE Transactions on Information Theory*, 46(6):1962–1973.

[Hochwald et al., 2005] Hochwald, B., Peel, C., and Swindlehurst, A. (2005). A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation. *IEEE Transactions on Communications*, 53(3).

[Hou and Kumar, 2010] Hou, I.-H. and Kumar, P. R. (2010). Utility-optimal scheduling in time-varying wireless networks with delay constraints. In *Proceedings of the eleventh ACM international symposium on Mobile ad hoc networking and computing*, page 31.

[Hou and Yang, 2011] Hou, X. and Yang, C. (2011). Codebook design and selection for multi-cell cooperative transmission limited feedback systems. In *IEEE Vehicular Technology Conference (VTC Spring)*.

[Inoue and Heath, 2009] Inoue, T. and Heath, R. (2009). Kerdock codes for limited feedback precoded MIMO systems. *IEEE Transactions on Signal Processing*, 57(9).

[Irmer et al., 2011] Irmer, R., Droste, H., Marsch, P., Grieger, M., Fettweis, G., Brueck, S., Mayer, H.-P., Thiele, L., and Jungnickel, V. (2011). Coordinated multipoint: concepts, performance, and field trial results. *IEEE Communications Magazine*, 49(2):102–111.

[James and Liebeck, 2001] James, G. and Liebeck, M. (2001). *Representations and Characters of Groups*. Cambridge University Press, 2nd edition.

[Jiang et al., 2012] Jiang, C., Wang, M., Shu, F., Wang, J., Sheng, W., and Chen, Q. (2012). Multi-User MIMO with Limited Feedback Using Alternating Codebooks. *IEEE Transactions on Communications*, 60(2).

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

[Joham et al., 2005] Joham, M., Utschick, W., and Nossek, J. (2005). Linear transmit processing in MIMO communications systems. *IEEE Trans. Signal Process.*, 53(8):2700–2712.

[Johansson and Xiao, 2006] Johansson, M. and Xiao, L. (2006). Cross-layer optimization of wireless networks using nonlinear column generation. *IEEE Transactions on Wireless Communications*, 5(2):435–445.

[Joung and Lee, 2007] Joung, J. and Lee, Y. (2007). Regularized channel diagonalization for multiuser mimo downlink using a modified mmse criterion. *IEEE Transactions on Signal Processing*, 55(4):1573–1579.

[Jungnickel and Tonchev, 1999] Jungnickel, D. and Tonchev, V. (1999). Decompositions of difference sets. *Journal of Algebra*, 217(1).

[Khalek et al., 2012] Khalek, A., Caramanis, C., and Heath, R. (2012). A cross-layer design for perceptual optimization of H.264/SVC with unequal error protection. *IEEE Journal on Selected Areas in Communications*, 30(7).

[Khisti and Wornell, 2010a] Khisti, A. and Wornell, G. (2010a). Secure transmission with multiple antennas I: The MISOME wiretap channel. *IEEE Trans. Inf. Theory*, 56(7):3088–3104.

[Khisti and Wornell, 2010b] Khisti, A. and Wornell, G. W. (2010b). Secure transmission with multiple antennas II: the MIMOME wiretap channel. *IEEE Transactions on Information Theory*, 56(11):5515–5532.

[Kim et al., 2011] Kim, J., Michalopoulos, D. S., and Schober, R. (2011). Diversity analysis of multi-user multi-relay networks. *IEEE Transactions on Wireless Communications*, 10(7):2380–2389.

[Kobayashi and Caire, 2006] Kobayashi, M. and Caire, G. (2006). An iterative water-filling algorithm for maximum weighted sum-rate of Gaussian MIMO-BC. *IEEE Journal on Selected Areas in Communications*, 24(8):1640–1646.

[Koyluoglu et al., 2012] Koyluoglu, O., Koksal, C., and El Gamal, H. (2012). On secrecy capacity scaling in wireless networks. *IEEE Transactions on Information Theory*, 58(5).

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

[Laneman et al., 2004] Laneman, J., Tse, D., and Wornell, G. (2004). Cooperative diversity in wireless networks: efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 50(12):3062–3080.

[Lee and Shin, 2011] Lee, N. and Shin, W. (2011). Adaptive feedback scheme on K-cell MISO interfering broadcast channel with limited feedback. *IEEE Transactions on Wireless Communications*, 10(2).

[Li and Liu, 2006] Li, G. and Liu, H. (2006). Downlink radio resource allocation for multi-cell OFDMA system. *IEEE Transactions on Wireless Communications*, 5(12).

[Li et al., 2010] Li, Q., Li, G., Lee, W., Lee, M., Mazzarese, D., Clerkx, B., and Li, Z. (2010). MIMO techniques in WiMAX and LTE: a feature overview. *IEEE Communications Magazine*, 48(5):86–92.

[Li et al., 2005] Li, Q., Lin, X., Zhang, J., Lindskog, C., Baum, K., and Wennstrom, M. (2005). Compact codebooks for transmit beamforming in closed-loop MIMO. Technical report, IEEE 802.16 Broadband Wireless Access Working Group.

[Liang and Goldsmith, 2007] Liang, Y. and Goldsmith, A. (2007). Adaptive channel reuse in cellular systems. In *IEEE International Conference on Communications*, pages 857–862. IEEE.

[Liang et al., 2009] Liang, Y., Poor, H., and Ying, L. (2009). Secrecy throughput of MANETs with malicious nodes. In *Proceedings IEEE Internation Symposium on Information Theory*.

[Liu et al., 2011] Liu, H., Song, Y., Li, D., Cai, L., Yang, H., Lu, D., and Wu, K. (2011). Scalable limited channel feedback for downlink coordinated multi-cell transmission. In *IEEE Vehicular Technology Conference (VTC Spring)*.

[Liu et al., 2012] Liu, L., Chen, R., Geirhofer, S., Sayana, K., Shi, Z., and Zhou, Y. (2012). Downlink MIMO in LTE-advanced: SU-MIMO vs. MU-MIMO. *IEEE Communications Magazine*, 50(2):140–147.

[Liu et al., 2006a] Liu, P., Berry, R., and Honig, M. (2006a). A fluid analysis of a utility-based wireless scheduling policy. *IEEE Transactions on Information Theory*, 52(7):2872–2889.

[Liu et al., 2006b] Liu, Q., Wang, X., and Giannakis, G. (2006b). A cross-layer scheduling algorithm with QoS support in wireless networks. *IEEE Transactions on Vehicular Technology*, 55(3):839–847.

[Liu et al., 2005] Liu, Q., Zhou, S., and Giannakis, G. (2005). Queuing with adaptive modulation and coding over wireless links: cross-Layer analysis and design. *IEEE Transactions on Wireless Communications*, 4(3):1142–1153.

[Love and Heath, 2005] Love, D. and Heath, R. (2005). Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Transactions on Information Theory*, 51(8).

[Love et al., 2003] Love, D., Heath, R., and Strohmer, T. (2003). Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Transactions on Information Theory*, 49(10):2735–2747.

[Marzetta, 2010] Marzetta, T. L. (2010). Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Transactions on Wireless Communications*, 9(11):3590–3600.

[Mondal et al., 2007] Mondal, B., Samanta, R., and Heath Jr., R. (2007). Congruent Voronoi tessellations from equiangular lines. *Applied and Computational Harmonic Analysis*, 23(2):254–258.

[Motorola, 2007] Motorola (2007). Long Term Evolution (LTE): A Technical Overview.

[Mukherjee et al., 2013] Mukherjee, A., Fakoorian, S., Huang, J., and Swindlehurst, A. (2013). Principles of physical-layer security in multiuser wireless networks: a survey. *submitted to IEEE Communications Surveys and Tutorials*.

[Mukherjee et al., 2010] Mukherjee, A., Fakoorian, S. A. A., Huang, J., and Swindlehurst, A. L. (2010). Principles of physical-layer security in multiuser wireless networks: Survey. arXiv:1011.3754.

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

[Neely et al., 2002] Neely, M., Modiano, E., and Rohrs, C. (2002). Power and server allocation in a multi-beam satellite with time varying channels. In *Proc. INFOCOM*, pages 1451–1460.

[Ng and Yu, 2007] Ng, T. and Yu, W. (2007). Joint optimization of relay strategies and resource allocations in cooperative cellular networks. *IEEE Journal on Selected Areas in Communications*, 25(2):328–339.

[Nguyen et al., 2009] Nguyen, V. K., Muharar, R., and Evans, J. S. (2009). Multiuser transmit beamforming via regularized channel inversion: A large system analysis. *Technical Report*. Available: http://cubinlab.ee.unimelb.edu.au/ rmuharar/doc/manuscriptrevRusdha22-11.pdf.

[Norris, 1997] Norris, J. (1997). *Markov Chains*. Cambridge University Press.

[Novlan et al., 2011] Novlan, T., Ghanti, R., Ghosh, A., and Andrews, J. (2011). Analytical evaluation of fractional frequency reuse for OFDMA cellular networks. *IEEE Transactions on Wireless Communications*, 10(12).

[O'Neill et al., 2008] O'Neill, D., Goldsmith, A., and Boyd, S. (2008). Wireless network utility maximization. In *IEEE Military Communications Conference.*, pages 1–8. IEEE.

[Palomar and Chiang, 2006] Palomar, D. and Chiang, M. (2006). A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications*, 24(8).

[Papailiopoulos et al., 2012] Papailiopoulos, D., Luo, J., Dimakis, A., Huang, C., and Li, J. (2012). Simple regenerating codes: network coding for cloud storage. In *Proceedings of IEEE INFOCOM*.

[Papandriopoulos et al., 2008] Papandriopoulos, J., Dey, S., and Evans, J. (2008). Optimal and distributed protocols for cross-layer design of physical and transport layers in MANETs. *IEEE/ACM Transactions on Networking*, 16(6):1392–1405.

[Park et al., 2005] Park, D., Seo, H., Kwon, H., and Lee, B. (2005). Wireless packet scheduling based on the cumulative distribution function of user transmission rates. *IEEE Transactions on Communications*, 53(11).

[Peel et al., 2005] Peel, C., Hochwald, B., and Swindlehurst, A. (2005). A vector-perturbation technique for near-capacity multiantenna multiuser communication part I: channel inversion and regularization. *IEEE Transactions on Communications*, 53(1):195–202.

[Ryan et al., 2009] Ryan, D., Clarkson, I., Collings, I., and Honig, M. (2009). QAM and PSK codebooks for limited feedback MIMO beamforming. *IEEE Transactions on Communications*, 57(4):1184–1196.

[Saquib et al., 20] Saquib, N., Hossain, E., and Kim, D. (20). Fractional frequency reuse for interference management in LTE-advanced hetnets. *IEEE Wireless Communications*, 2.

[Scott and Grassl, 2010] Scott, A. J. and Grassl, M. (2010). Symmetric informationally complete positive-operator-valued measures: A new computer study. *Journal of Mathematical Physics*, 51(4):042203.

[Seifi et al., 2011] Seifi, N., Matthaiou, M., and Viberg, M. (2011). Coordinated user scheduling in the multi-cell MIMO downlink. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2840–2843. IEEE.

[Sendonaris et al., 2003] Sendonaris, A., Erkip, E., and Aazhang, B. (2003). User cooperation diversity. part I. system description. *IEEE Transactions on Communications*, 51(11).

[Shamai and Zaidel, 2001] Shamai, S. and Zaidel, B. (2001). Enhancing the cellular downlink capacity via co-processing at the transmitting end. In *IEEE Vehicular Technology Conference (VTC Spring)*.

[Sharif and Hassibi, 2007] Sharif, M. and Hassibi, B. (2007). Delay considerations for opportunistic scheduling in broadcast fading channels. *IEEE Transactions on Wireless Communications*, 6(9):3353–3363.

[Shiu et al., 2011] Shiu, Y.-S., Chang, S., and Wu, H.-C. (2011). Physical layer security in wireless networks: a tutorial. *IEEE Wireless Communications*, 18(2).

[Shokrollahi et al., 2001] Shokrollahi, A., Hassibi, B., Hochwald, B., and Sweldens, W. (2001). Representation theory for high-rate multiple-antenna code design. *IEEE Transactions on Information Theory*, 47(6):2335–2367.

[Small Cell Forum, 2012] Small Cell Forum (2012). Small cells - what's the big idea? Femtocells are expanding beyond the home. Technical report.

[So-In et al., 2010] So-In, C., Jain, R., and Al-Tamimi, A.-K. (2010). A scheduler for unsolicited grant service (UGS) in IEEE 802.16e mobile WiMAX networks. *IEEE Systems Journal*, 4(4):487–494.

[So-In et al., 2009] So-In, C., Jain, R., and Tamimi, A.-K. (2009). Scheduling in IEEE 802.16e mobile WiMAX networks: Key issues and a survey. *IEEE Journal on Selected Areas in Communications*, 27(2):156–171.

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

[Strohmer and Heath, 2003] Strohmer, T. and Heath, R. (2003). Grassmannian frames with applications to coding and communication. *Applied and Computational Harmonic Analysis*, 14(3):257–275.

[Su et al., 2011] Su, D., Hou, X., and Yang, C. (2011). Quantization based on per-cell codebook in cooperative multi-cell systems. In *IEEE Wireless Communications and Networking Conference (WCNC)*.

[Sung and Collings, 2010] Sung, C. K. and Collings, I. (2010). Cooperative transmission with decode-and-forward MIMO relaying in multiuser relay networks. In *Proc. IEEE Int. Conf. on Comm. (ICC)*.

[Sung et al., 2009] Sung, H., Lee, S.-R., and Lee, I. (2009). Generalized channel inversion methods for multiuser mimo systems. *IEEE Transactions on Communications*, 57(11):3489–3499.

[Suzuki et al., 2012] Suzuki, H., Collings, I., Hayman, D., Pathikulangara, J., Chen, Z., and Kendall, R. (2012). Large-scale multiple antenna fixed wireless systems for rural areas. In *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*.

[Tang and Zhang, 2007] Tang, J. and Zhang, X. (2007). Cross-layer resource allocation over wireless relay networks for quality of service provisioning. *IEEE Journal on Selected Areas in Communications*, 25(4):645–656.

[Telatar, 1999] Telatar, I. (1999). Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6).

[Telatar and Gallager, 1995] Telatar, I. and Gallager, R. (1995). Combining queueing theory with information theory for multiaccess. *IEEE Journal on Selected Areas in Communications*, 13(6).

[Tropp et al., 2005] Tropp, J., Dhillon, I., Heath, R., and Strohmer, T. (2005). Designing structured tight frames via an alternating projection method. *IEEE Transactions on Information Theory*, 51(1).

[Vasudevan et al., 2010] Vasudevan, S., Goeckel, D., and Towsley, D. (2010). Security-capacity trade-off in large wireless networks using keyless secrecy. In *Proc. ACM Int. Symp. Mobile Ad Hoc Networking and Computing (MobiHoc)*.

[Wiesel et al., 2008] Wiesel, A., Eldar, Y., and Shamai, S. (2008). Zero-forcing precoding and generalized inverses. *IEEE Transactions on Signal Processing*, 56(9).

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

[Yang et al., 2011] Yang, N., Elkashlan, M., and Yuan, J. (2011). Impact of opportunistic scheduling on cooperative dual-hop relay networks. *IEEE Transactions on Communications*, 59(3):689–694.

[Ye et al., 2013] Ye, Q., Rong, B., Chen, Y., Al-Shalash, M., Caramanis, C., and Andrews, J. G. (2013). User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, 12(6).

[Yeh and Cohen, 2003] Yeh, E. and Cohen, A. (2003). A fundamental cross-layer approach to uplink resource allocation. In *IEEE Military Communications Conference.*, pages 699–704. IEEE.

[Yoo and Goldsmith, 2006] Yoo, T. and Goldsmith, A. (2006). On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *IEEE J. Sel. Areas Commun.*, 24(3):528–541.

[Zalesky, 2009] Zalesky, A. (2009). To burst or circuit switch? *IEEE/ACM Transactions on Networking*, 17(1):305–318.

[Zhang et al., 2011] Zhang, H., Venturino, L., Prasad, N., Li, P., Rangarajan, S., and Wang, X. (2011). Weighted sum-rate maximization in multi-cell networks via coordinated scheduling and discrete power control. *IEEE Journal on Selected Areas in Communications*, 29(6):1214–1224.

[Zhang and Andrews, 2010] Zhang, J. and Andrews, J. G. (2010). Adaptive spatial intercell interference cancellation in multicell wireless networks. *IEEE Journal on Selected Areas in Communications*, 28(9):1455–1468.

[Zhou et al., 2011] Zhou, S., Gong, J., and Niu, Z. (2011). Distributed adaptation of quantized feedback for downlink network MIMO systems. *IEEE Transactions on Wireless Communications*, 10(1).