Thesis for the degree of Doctor of Philosophy

Base Station Coordination in Multicell MIMO Networks

Nima Seifi

CHALMERS

Communication Systems and Information Theory Group
Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2012

Base Station Coordination in Multicell MIMO Networks
Nima Seifi
ISBN 978-91-7385-649-2

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 3330
ISSN 0346-718X

Communication Systems and Information Theory Group
Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: + 46 (0)31-772 1000

Front Cover:
Illustration of Base Station Coordination in a Two-Cell Network

This thesis has been prepared using LaTeX.
Printed by Chalmers Reproservice,
Gothenburg, Sweden, 2012.

To my friends

My heart which in this desert land,
Hath wandered far to understand,
Yet, ever so simple, a blade of grass,
Doth lie beyond its reach- alas.

And though within this heart of mine,
A thousand suns have come to shine,
No closer am I to understand,
The truth within a grain of sand.

Avicenna, Persian Philosopher, 980-1037 A.D.

Peom translation by: Babak Hassibi, Stanford, 1995

Abstract

The use of multi-antenna technology, also referred to as multiple-input multiple-output (MIMO), has been shown to improve both the achievable data rates and the link reliability in single-cell wireless systems without a need for extra power or bandwidth. The promised gains of MIMO techniques are, however, severely degraded in a multicell environment due to the presence of intercell interference, especially for users at the cell edge. One efficient technique to combat intercell interference is via exploiting coordination among multiple base stations, which is known as multicell processing or simply base station coordination.

This thesis investigates the design and the performance of practically implementable base station coordination schemes. The main contribution of this thesis is to formally study different types of coordination, to develop analytical tools for their performance evaluation, and to propose simple algorithms for their implementation.

First, we focus on the most complex form of coordination, namely the *network MIMO*. In this scheme all coordinating base stations share the data and the channel state information of all users, and act as a single distributed multi-antenna transmitter to serve them. We develop an analytical framework to facilitate the ergodic rate analysis of such a system under linear precoding. We also propose a simple scheduling algorithm, which only requires the knowledge of long-term channel statistics.

In the next stage, we consider a simpler form of coordination in which the data of each user is served only by one base station. The scheduling and beamforming design, however, can be shared among the coordinating base stations. For this scheme, we propose a low-complexity joint user scheduling and beamforming strategy selection which requires a limited level of inter-base station information exchange, while providing significant performance improvement over non-coordinated systems.

Finally, we investigate the effect of the antenna elevation tuning parameter, referred to as *antenna tilt*, on the performance of multicell multiple-input single-output (MISO) systems. We propose a framework in which multiple base stations jointly adjust their tilt angles based on the location of the scheduled users to maximize their sum throughput. We also provide an analytical expression for the sum throughput, enabling the decentralized implementation of the proposed scheme at each base station.

Keywords: Antenna tilt, base station coordination, beamforming, intercell interference, scheduling.

Base Station Coordination in Multicell MIMO Networks
NIMA SEIFI
Communication Systems and Information Theory Group
Department of Signals and Systems

List of Publications

This thesis is based on the following publications:

Paper A
N. Seifi, M. Viberg, R. W. Heath Jr., J. Zhang, and M. Coldrey, "Multimode Transmission in Network MIMO Downlink with Incomplete CSI", *EURASIP J. Adv. Signal Process., special issue on Cooperative MIMO Multicell Networks*, 2011.

Paper B
N. Seifi, M. Matthaiou, M. Viberg, and M. Coldrey, "On the Achievable Ergodic Rate of Network MIMO Systems With Imperfect CSI", submitted to *IEEE Wireless Commun. Letter*, Feb. 2012.

Paper C
N. Seifi, J. Zhang, M. Viberg, and R. W. Heath Jr., "Joint Scheduling and Intercell Interference Management in Multicell MISO Networks", submitted to *IEEE Trans. Signal Process.*, Aug. 2011.

Paper D
N. Seifi, M. Matthaiou, and M. Viberg, "Coordinated User Scheduling in the Multicell MIMO Downlink", in Proc. of *IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011.

Paper E
N. Seifi, M. Coldrey, and M. Viberg, "Throughput Optimization in Multicell MISO Network via Coordinated User-Specific Tilting", submitted to *IEEE Commun. Letter*, Jan. 2012.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Mats Viberg, for giving me the opportunity to become a PhD student at the department of Signals and Systems, and the freedom to work on so many interesting research problems. Your knowledge, guidance and supportive attitude have been fundamental to my growth in every aspect during my stay at Chalmers. I always became more enthusiastic and committed to my research after every single research meeting I had with Mats. I'm also very grateful to the head of the group, Professor Erik Ström, for creating such a friendly atmosphere in the group and for all I have learnt from him during the years I have been his teaching assistant in Wireless Communication course.

I'd also like to mention my deepest gratitude to Professor Robert W. Heath Jr. for giving me the opportunity to visit the Wireless Networking and Communication Group (WNCG) at The University of Texas at Austin. Your sharp mind in research discussions and your wonderful research leadership skills have always been a source of inspiration for me. I'm also grateful to Professor Jun Zhang, who was a post-doc in WNCG during my visit, for reading the rough draft of my papers and patiently iterating them toward a high-quality final version. I'm also thankful to all the PhD students in WNCG for their hospitality and for all the fun discussions we had together. Without a doubt, visiting WNCG has made a mark in my PhD studies.

I have mainly worked within the Chase-Charmant MIMO Systems project for which I would like to thank the project partners Ericsson and Qamcom. I would specially like to thank Dr. Mikael Coldrey for his supportive attitude, and for following my work carefully over the last years. This gratitude also goes to Professor Tony Ottosson, my supervisor before the Licentiate degree, and Dr. Andreas Wolfgang, who took me under his wing from the beginning of my graduate studies. I would also like to thank Dr. Michail Matthaiou for our collaboration over the last year. At the same time, I thank the former Chase manager, Ingmar Karlsson, the current manager Staffan Sjödin, and the director of Charmant, Professor Mats Viberg, for creating and supporting such a great research environment.

I am also deeply indebted to current and past members of the communication systems group – too many to name here, for making it a great place to work. It has been a wonderful experience to be one of you. I'm also thankful to my friends outside the communication systems group – also too many to name here, for all the great moments we have experienced together.

I would also like to use the opportunity to send my gratitude to my dear Iranian friends in Austin. Special thanks goes to Babak Fallah Azad for providing me with all kinds of information about Austin before my departure, for picking me up at the airport, for helping me out to furnish my apartment, and for buying me soup and medicine when I was heavily ill. This special thanks also goes to Yousof Mortazavi, Ardavan Pedram, Amir Shahed Mehrtash for all the salsa nights we attended together, for all the free drives they gave me, and for all the great times we spent together. Without you this visit would never become as amazing as it did now.

On the administrative side, I would like to express my appreciation to Agneta, Natasha, and Madeleine. I would also like to thank Lars Börjesson for his computer-related support. My gratitude also goes to the beautiful personnel at Café Bulten for making the best Latte (with skimmed milk!) in this town and for teaching me Swedish!

Finally, I'd like to express my sincerest gratitude to my family for their constant support throughout the ups and downs of my studies. Without you I couldn't have made it to this point! My special thanks goes to the best friend of all times, Hooman, which also happens to be my brother, for being such a good company in all the sad and happy moments in my life.

Nima Seifi
Gothenburg, January 2012

Contents

Acronyms

3G:	Third Generation
3GPP:	Third Generation Partnership Project
AWGN:	Additive White Gaussian Noise
BC:	Broadcast Channel
BD:	Block Diagonalization
BF:	Beam-Forming
bps:	bits per second
BS:	Base Station
CDF:	Cumulative Distribution Function
CSB:	Coordinated Scheduling/Beamforming
CSI:	Channel State Information
DPC:	Dirty Paper Coding
FB:	Feedback
FDD:	Frequency Division Duplex
FDMA:	Frequency Division Multiple Access
LOS:	Line-Of-Sight
MAC:	Multiple Access Channel
MIMO:	Multiple-Input Multiple-Output
MISO:	Multiple-Input Single-Output
MMSE:	Minimum Mean Square Error
MMT:	Multi-Mode Transmission
MU:	Multi-User

NLOS:	Non-Line-Of-Sight
OFDM:	Orthogonal Frequency Division Multiplexing
PDF:	Probability Density Function
PF:	Proportional Fair
QoS:	Quality of Service
RR:	Round Robin
SDMA:	Space Division Multiple Access
SINR:	Signal-to-Interference-plus-Noise Ratio
SISO:	Single-Input Single-Output
SM:	Spatial Multiplexing
SNR:	Signal-to-Noise Ratio
SU:	Single-User
SVD:	Singular Value Decomposition
TDD:	Time Division Duplex
TDMA:	Time Division Multiple Access
WF:	Water-Filling
ZF:	Zero-Forcing

Part I

Introduction

Chapter 1

Overview

The convergence between mobile and data access services has caused an increasing demand for high data rate wireless communications. The availability of affordable notebooks, tablet computers, smartphones, etc. as well as a wide range of services including web browsing, streaming, and interactive file transfer has resulted in a significant growth in the mobile data traffic recently. This growth is continuing rapidly such that a typical subscriber is expected to consume 1 gigabyte of data per month by 2014, while the today's average figure is about a few hundred megabytes per month [1]. So, eventually the current 3G network will not be able to support the traffic demand and, hence, a more advanced and efficient wireless technology is needed to provide the required services. This technology should also provide a seamless experience with a guaranteed minimum quality of service to all the users irrespective of their location in the network.

To this end, the *third generation partnership project* (3GPP) has been developing a new mobile communication standard, referred to as *long term evolution* (LTE). Its aim is to provide true 4G broadband mobile access, i.e., to fulfill the international mobile telecommunication advanced (IMT-Advanced) requirements as defined by the international telecommunication union (ITU)–such as peak data rate up to 1 Gbits/s [2]. The first release of LTE (release 8) is labeled as 3.9G (beyond 3G but pre-4G) as it does not meet the IMT-Advanced requirements for 4G. However, Release 10 of LTE, also referred to as LTE-Advanced, is considered as a true 4G evolution step. The first commercial LTE networks were launched in Sweden and Norway in December 2009 followed by the United States and Japan in 2010.

The use of orthogonal frequency division multiplexing (OFDM) and multi-antenna technology, commonly referred to as multiple-input multiple-output (MIMO), has enabled the emerging wireless standard (including LTE) to achieve a significant spectral efficiency within one cell. In a multicell network, however, the presence of intercell interference, caused by the transmission of neighboring cells on the same time-frequency resource block, prevents these technologies to approach the theoretical rates of multicell networks. Intercell interference especially affects the performance of users at the cell boundaries. Compared to users in cell-center region, users in the cell-edge region are subject to a weaker desired signal from their *serving* base station, while experiencing a stronger intercell interference from the neighboring cells.

One of the promising approaches to combat intercell interference is to exploit coordination among multiple base stations, a.k.a. multicell processing. This technique is currently under investigation in the emerging wireless standards such as LTE-Advanced, under the name of *coordinated multipoint transmission* (CoMP). The main idea of CoMP is to either mitigate or exploit the intercell interference in order to improve the cell-edge and average data rate. Under ideal conditions, in which the channel state information (CSI) and the data of all users is fully shared among all the base stations, coordinated multicell transmission have been shown to provide a remarkable throughput gain compared to the conventional non-coordinated systems [3].

Implementation of base station coordination in practice faces some major challenges concerning complexity and overhead. Users' CSI at all the coordinating base stations needs to be obtained through some form of training, channel estimation, and feedback. The acquired CSI is usually imperfect as the estimation and feedback incurs error. Furthermore, the acquired CSI and the data needs to be shared among the coordinating base stations through finite-capacity backhaul links which are subject to both error and delay. These challenges prevent the base station coordination to reach its promised theoretical gains over the conventional non-coordinated systems [4, 5].

In this thesis, we investigate different methods of coordinated transmission from multiple base stations. The aim is to reduce the complexity and overhead, while keeping the performance as close to the theoretical limit as possible. The general methodology followed throughout this thesis is the development of mathematical tools that can be used to simplify the analysis of coordinated schemes. In addition, new algorithms are also proposed to reduce the complexity and overhead for different coordinated transmission schemes.

Research Project

The research project in this thesis is part of the MIMO system projects in Chalmers strategic research center on microwave antenna systems (CHARMANT) and Chalmers antenna system excellence center (CHASE) at Chalmers University of Technology. The project has been done in collaboration with Ericsson AB and Qamcom AB.

Outline of the Thesis

The thesis is divided into two parts. Part I gives a quick introduction to the main topic of the thesis to facilitate the understanding of the included contributions in Part II. This part is organized as follows. In Chapter 2, the traditional single-antenna mutlicell wireless networks are discussed, and several methods for intercell interference mitigation in such networks are presented. Chapter 3 gives an overview of multi-antenna transmission techniques in multicell wireless networks and discusses the motivation behind multicell coordination. In Chapter 4, different multicell coordination strategies and the corresponding practical challenges are described. Finally, Chapter 5 briefly describes the purpose and the contributions included in Part II of the thesis. Some open problems in the field as well as a list of related contributions not included in this thesis are also presented in this chapter.

Chapter 2

Conventional Single-Antenna Cellular Networks

This chapter gives a brief introduction to the operation of conventional cellular networks consisting of single-antenna transmitters and receivers. The concepts of fading and intercell interference as two fundamental challenges in these networks are discussed and traditional methods to deal with them are presented. The purpose of the chapter is to provide sufficient background to motivate the use of multiple antennas and base station coordination in cellular networks.

2.1 Wireless Channel

In wireless systems, transmission of data from a transmitter to a receiver is performed via propagation of electromagnetic waves over an unguided environment, referred to as the *wireless channel*. The propagated waves are reflected, scattered, and diffracted by walls, buildings, trees, and other objects as they travel from the transmitter towards the receiver. Such a propagation results in the arrival of multiple replicas of the transmitted signal at the receiver, each with different characteristics (delay, gain, phase, etc.), as illustrated in Fig. 2.1. Hence, the wireless channel is sometimes referred to as a *multipath channel*. The detailed description of this propagation can be obtained through solving Maxwell's equations [6]. However, this description would be very difficult in general, as many of the required parameters are not available. Furthermore, the wireless channel varies with time as either the transmitter or the receiver, or both change position. To deal with this

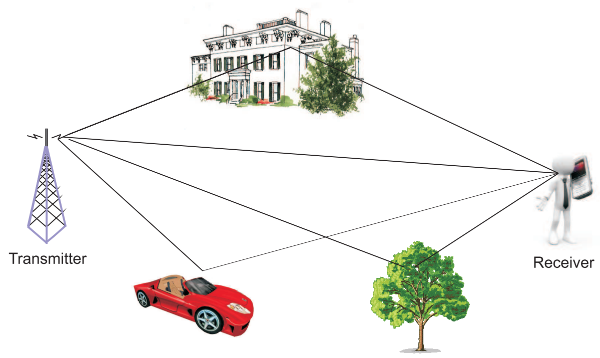


Figure 2.1: A schematic illustration of multipath propagation in wireless channel.

difficulty, statistical models have been developed to characterize the time-varying behavior of the wireless channel (see [7] and the references therein). In these models, the wireless channel variation is described at two different time-scales as follows:

- **Large-scale fading:** Slow channel variations as the distance between the transmitter and the receiver changes significantly over a time-scale on the order of tens of seconds [8]. The variations in large-scale fading are mainly due to pathloss and shadowing. Pathloss describes the decay in the received signal power owing to the distance between the transmitter and the receiver. Shadowing is the attenuation in the received signal power as a result of absorption, reflection, scattering, and diffraction of the transmitted signal through large obstacles between the transmitter and the receiver.

- **Small-scale fading:** Fast channel variation due to small changes in the relative spatial position of the transmitter and the receiver over a time-scale on the order of a few milliseconds [8]. Small-scale fading channel variations result from the constructive and destructive addition of multipath signal components.

It is normally assumed that the large-scale fading is known for all the users in the network. This is a reasonable assumption as the large-scale fading changes very slowly so that it can be accurately measured by each user and fed back to the base station. We next provide a short review of different phenomena in a wireless channel. For more detailed information, the interested reader is referred to traditional wireless communication text books such as [6, 7].

2.1.1 Pathloss

Pathloss in wireless channel refers to the received signal power attenuation as a result of distance between the transmitter and the receiver. Pathloss in linear scale is defined as the ratio of the transmitted power to the received power, i.e.,

$$\mathrm{PL} = \frac{P_{\mathsf{t}}}{P_{\mathsf{r}}}, \tag{2.1}$$

where P_{t} is the transmitted power, while P_{r} denotes the received power. The path gain is defined as the inverse of pathloss, i.e., $\mathrm{PG} = 1/\mathrm{PL}$. Note that the pathloss is greater than 1 and hence, the path gain is always less than 1. The pathloss in free space is given by the Friis formula as

$$\mathrm{PL} = \frac{\lambda G}{(4\pi d)^2}, \tag{2.2}$$

where λ is the wavelength, G is the product of antenna gains at both the transmitter and the receiver, and d is the distance between the transmitter and the receiver. It is observed that free space pathloss falls off in inverse proportion to the square of the distance. This model is only valid when there is one direct path (a.k.a. line-of-sight) between the transmitter and the receiver. To capture the pathloss in a cellular environments with multiple signal paths between the transmitter and the receiver, several deterministic and statistical models have been developed [7, Chapter 2.]. A simplified pathloss model commonly used in wireless system design is given by

$$\mathrm{PL} = \xi d^{-\upsilon}. \tag{2.3}$$

Here, ξ is a constant that depends on the antenna characteristics and the average channel attenuation, and υ is the pathloss exponent. The value of υ changes between 2 and 6, depending on the propagation environment.

2.1.2 Shadow Fading

In a multipath wireless channel, the transmitted signal will experience blockage from large objects and reflecting surfaces in the signal path. This results in random variation in the received signal power at a given distance. As the size, shape, and location of these objects are usually unknown, an easy way to model these variations is via statistical methods. The most common model is the log-normal shadowing. This model has been empirically confirmed to accurately represent the received signal variations both in outdoor and indoor environments [7, Chapter 2]. The probability density function (PDF) of a log-normal random variable x is given by

$$p(x) = \frac{10}{x\sigma_{x_{\mathsf{dB}}} \ln(10)\sqrt{2\pi}} e^{-\frac{(10\log_{10} x - \mu_{x_{\mathsf{dB}}})^2}{2\sigma^2_{x_{\mathsf{dB}}}}}, x > 0. \tag{2.4}$$

Here, $\mu_{x_{\mathsf{dB}}}$ and $\sigma_{x_{\mathsf{dB}}}$ denote the mean and the standard deviation of x in dB scale, which is computed as $x_{\mathsf{dB}} = 10\log_{10}(x)$.

2.1.3 Small-Scale Fading

Small-scale fading refers to the microscopic changes in the signal amplitude due to constructive and destructive addition of multipath components over a time-scale on the order of a few milliseconds. These small variations, which occur on top of the variations introduced by shadowing, result from small changes in the spatial position between the transmitter and the receiver in the range of a few wavelengths. The impact of multipath fading on the received signal depends on the spread of delays associated with multipath components. Let τ_i denote the delay of the i-th multipath component, then the delay spread of the channel is defined as $T_{\mathsf{ds}} = \max_{i,j}(\tau_i - \tau_j)$. If this delay spread is much smaller than the inverse of the signal bandwidth, then the multipath components are considered as nonresolvable, leading to a single-tap channel. On the other hand, if the delay spread is comparable or larger than the inverse of the signal bandwidth, the multipath components become resolvable, giving rise to a multi-tap channel. Such a channel causes a form of distortion in the received signal, known as the *inter-symbol interference*, owing to the delayed replicas of the past symbols interfering with the current one. Methods to combat inter-symbol interference will be discussed later in this chapter. First, we describe three commonly used distributions for single-tap fading channels.

Rayleigh Fading

This is a valid model for scenarios where there is no line-of-sight path between the transmitter and the receiver, and there is a large number of independent scattered signal components. Under these conditions, using the Central Limit Theorem, the channel gain is modeled as a zero-mean complex-valued Gaussian process $z \sim \mathcal{CN}(0, \sigma^2)$. The PDF of the envelope of the channel gain, i.e., $r = |z|$, is given by

$$p(r) = \frac{2r}{\sigma^2} e^{-\frac{r^2}{\sigma^2}}, r > 0, \tag{2.5}$$

where $\sigma^2 = \mathbb{E}[r^2]$ is the average received power.

Rician Fading

This model is used when there is a fixed line-of-sight path between the transmitter and the receiver. The received signal is composed of a large number of complex Gaussian components plus a line-of-sight component. In this case, the signal envelope is well modeled with a so-called Rician distribution. In Rician fading, the severity of fading model is measured with a fading parameter K, which is the ratio of the power of the line-of-sight component to the power of the other multipath components. The PDF of the signal envelope with Rician distribution is given by

$$p(r) = \frac{2r(K+1)}{\sigma^2} e^{(-K - \frac{(K+1)r^2}{\sigma^2})} I_0 \left(2r \sqrt{\frac{K(K+1)}{\sigma^2}} \right), r > 0, \tag{2.6}$$

where $\sigma^2 = \mathbb{E}[r^2]$ is the average received power and

$$I_0(r) = \frac{1}{2\pi} \int_0^{2\pi} e^{-x \cos\theta} d\theta \tag{2.7}$$

denotes the modified Bessel function of zero-th order.

Nakagami Fading

Rayleigh and Rician fading are obtained based on mathematical modeling of the underlying physical phenomena in the wireless channel. Hence, some experimental data does not fit with either of these distributions. Nakagami fading is a more general fading distribution that could be adjusted to fit with empirical measurements. The PDF of a signal envelope with Nakagami distribution is given by

$$p(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\sigma^{2m}} e^{-\frac{mr^2}{\sigma^2}}, r > 0, \tag{2.8}$$

where $\sigma^2 = \mathbb{E}[r^2]$ is the average received power, $\Gamma(m)$ is the Gamma function, and m determines the severity of fading. For $m = 1$, the Nakagami distribution reduces to the Rayleigh distribution, while for $m = (K+1)^2/(2K+1)$ it reduces to the Rician distribution with parameter K.

2.1.4 Coherence Bandwidth and Frequency Selectivity

The behavior of small-scale fading in the frequency domain can be characterized through a parameter called *coherence bandwidth*. This parameter is a statistical measure for the range of frequencies over which the channel is considered “flat”, i.e., constructive and destructive addition of multipath signals does not vary over this frequency range. The coherence bandwidth W_{c} of a channel is related to the delay spread T_{ds} of that channel as $W_{\mathsf{c}} \propto 1/T_{\mathsf{ds}}$. When the signal bandwidth is smaller than the channel coherence bandwidth, the channel behaves like an ideal filter with a constant gain over the whole signal bandwidth. Such a channel is commonly known as a flat-fading channel and the corresponding transmission over this channel is classified as a *narrowband* transmission. To enable higher data rate transmission, one technique is to increase the bandwidth of the transmitted signal. However, when the bandwidth of the transmitted signal becomes larger than the coherence bandwidth of the channel, different parts of the signal bandwidth will be treated differently by the channel, giving rise to a so-called frequency-selective fading channel. The corresponding transmission over such a channel is considered as a *wideband* transmission. Noting the coherence bandwidth relation with the delay spread, it follows that frequency-selective fading results in inter-symbol interference discussed in Section 2.1.3.

To combat this problem, most of the emerging wireless standards such as LTE, have adopted OFDM. In this technique, a wideband high-rate transmission is divided into many parallel narrowband low-rate transmission, thereby eliminating the inter-symbol interference [9]. Hence, it is common to focus on a narrowband transmission over a frequency-flat fading channel and to assume that the wideband transmission can be treated by applying the provided analysis to each of the narrowband frequency blocks in an OFDM symbol.

2.1.5 Coherence Time and Time Selectivity

The small-scale fading variations in the time domain are caused by the relative mobility of the transmitter and the receiver and are characterized through a parameter called *coherence time*. This parameter represents the time interval over which the small-scale channel variation is negligible as the relative position of the transmitter and the receiver changes. The coherence time T_c of the channel is connected to the Doppler spread f_D as $T_\mathsf{c} \propto 1/f_\mathsf{D}$, where f_D is equal to the maximum Doppler shift and is given by

$$f_\mathsf{D} = \frac{v}{\lambda}. \tag{2.9}$$

Here, v denotes the relative speed of the transmitter and the receiver and λ is the carrier wavelength. The shorter the coherence time, the faster the channel changes with time. The channel is considered as *time-selective* when the coherence time is much smaller than the symbol duration T_s, i.e., $T_\mathsf{c} \ll T_\mathsf{s}$, and is considered as *time-flat* otherwise.

2.2 Cellular Architecture

The basic idea behind the cellular network is to reuse the same frequency band at different locations of the coverage area. This idea relies on the fact that the signal power decays with distance. In a conventional single-antenna cellular architecture, the coverage area is divided into non-overlapping cells

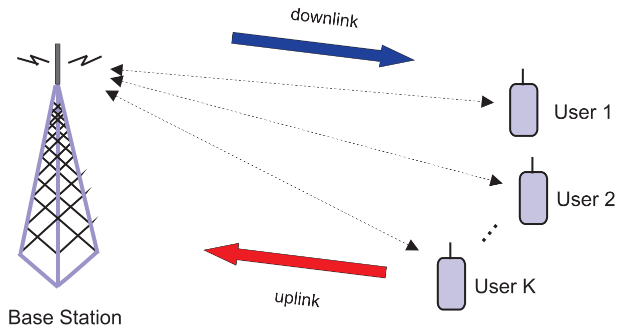


Figure 2.2: An illustration of downlink and uplink in a single cell multiuser system.

which operate independently, i.e., with no cell cooperation. Each cell has a single-antenna base station in the center and multiple single-antenna users distributed randomly over the cell area. Two typical scenarios at each cell are downlink transmission, which indicates the transmission from base station to users, and uplink transmission, which represent the transmission from users towards the base station. This is illustrated in Fig. 2.2.

At each time-slot, the base station in each cell communicates only with one user in that cell. As a result, there is no *intracell interference*. However, since the wireless channel is a shared medium and cells operate independently, the transmissions in the neighboring cells on the same frequency band act as interference to each other. Such an uncoordinated interference is known as *intercell interference*. The intercell interference in the downlink is explained for a three-cell network in Fig. 2.3.

Let $R_k(t)$ denote the instantaneous rate of user k in bits/s/Hz at time-slot t. This instantaneous rate is a function of signal-to-interference-plus-noise ratio (SINR) defined as

$$\mathsf{SINR}_k = \frac{P_{\mathsf{ds}}(t)}{P_{\mathsf{noise}}(t) + P_{\mathsf{IUI}}(t) + P_{\mathsf{ICI}}(t)}. \tag{2.10}$$

Here, $P_{\mathsf{ds}}(t)$, $P_{\mathsf{IUI}}(t)$, $P_{\mathsf{ICI}}(t)$, and $P_{\mathsf{noise}}(t)$ denote the power of the desired signal, the inter-user interference (or the intracell interference), the intercell interference, and the noise, respectively, at time-slot t. One commonly used instantaneous rate function is the so-called Shannon capacity formula given by [10]

$$R_k(t) = \log_2(1 + \mathsf{SINR}_k(t)). \tag{2.11}$$

Note that in (2.11), a sub-optimal receiver is assumed which treats the interference as Gaussian noise. In a conventional single-antenna cellular network, $P_{\mathsf{IUI}}(t)$ is zero since there is only one user communicating with

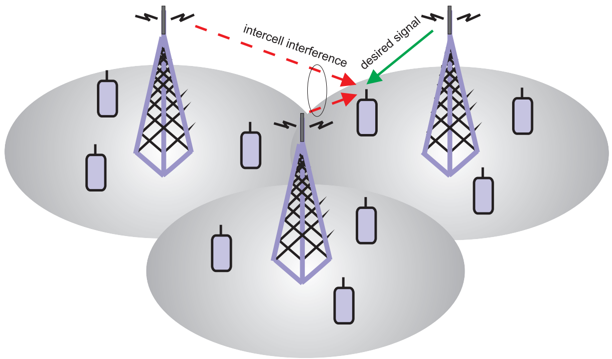


Figure 2.3: Schematic illustration of intercell interference in a 3-cell network.

the base station at each time-slot. However, $P_{\mathsf{ICI}}(t)$ is non-zero and could have large values, especially for users at the cell edge. In fact, in a realistic propagation scenario, where users are subject to distance-dependent pathloss and shadowing, cell-edge users experience a weaker desired signal power and a stronger intercell interference power, compared to users close to base stations, resulting in low SINR values for these users. Therefore, intercell interference is mainly a problem for the users at the cell edge and less of a problem for the cell-center users. This calls for efficient intercell interference mitigation techniques.

2.3 Intercell Interference Mitigation Techniques

As mentioned in the previous section, intercell interference could be a severe problem in cellular networks especially for the users close to the cell edge. In this section, we discuss some important techniques such as frequency reuse, cell sectoring, and cell densification, to mitigate this deteriorating phenomenon.

2.3.1 Frequency Reuse

Frequency reuse is one of the traditional techniques to control the level of intercell interference experienced by different users in the cell. In one way of implementing frequency reuse, known as *static* frequency reuse, the available bandwidth is divided into a number of disjoint subchannels each assigned to one of the neighboring cells. In this way, the interfering cells for any given

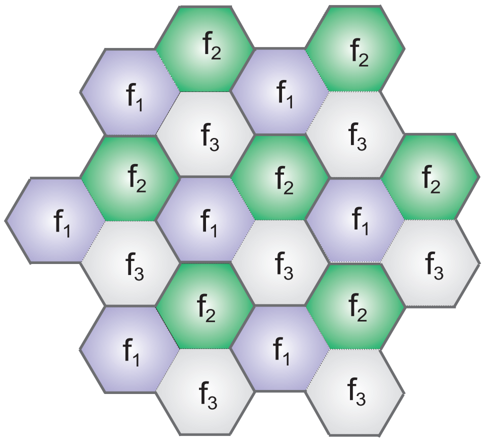


Figure 2.4: An example of static frequency reuse with reuse factor 3.

cell can be shifted away, thereby lowering the level of intercell interference power experienced by the users in that cell. This is illustrated in Fig. 2.4 for a frequency reuse factor of 3, where each of the three adjacent cells is assigned a different subchannel. The interfering cells are separated away by one-cell spacing. Note that, although static frequency reuse increases the SINR, it does not necessarily increase the achievable rate of all users. In this scenario, the instantaneous rate expression in (2.11) should be multiplied by $1/F$, where F is the reuse factor. For $F > 1$, some users, who gain less from intercell interference mitigation, might experience a lower data rate compared to universal frequency reuse ($F = 1$). This issue has motivated the proposal of *fractional* frequency reuse schemes in order to shift the effective reuse factor towards 1 [11].

Fractional frequency reuse is based on the idea of reuse partitioning [12], where a lower reuse factor is used in high SINR regions in the cell, while a higher reuse factor is used in low SINR regions. With universal frequency reuse, the highest throughput is provided to users in the cell-center who are experiencing a high SINR. With frequency reuse larger than 1, cell-edge users will experience the highest relative increase in the throughput as a result of intercell interference mitigation, compared to other users in the cell. The fractional frequency reuse scheme is actually a combination thereof where, for example, frequency reuse 1 is used for cell-center users while a frequency reuse larger than 1 is used for the cell-edge users. This is shown in Fig. 2.5, where the total frequency band is divided into four subchannels, namely f_1, f_2, f_3, and f_4. The subchannel f_1 is used in all the cells to serve users with high SINR. A frequency reuse of 3 is implemented at the cell edge on the remaining three subchannels. Note that in fractional

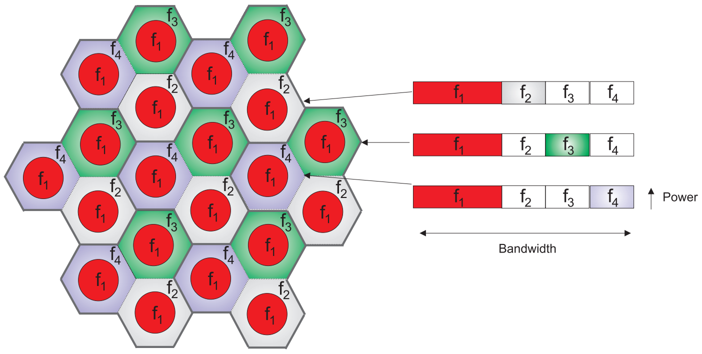


Figure 2.5: An example of fractional frequency reuse.

frequency reuse, the subchannels used at the edge of a given cell, are left empty in the neighboring cells. This results in the effective frequency reuse factor to be still less than 1, as it is not possible to transmit over the whole bandwidth in each cell. Soft frequency reuse is an attempt to address this problem [11].

In soft frequency reuse, it is possible to transmit over the whole system bandwidth, but with a non-uniform power spectrum. Figure 2.6 shows the power spectrum assignments in different cells of a system with soft frequency reuse. Notice that in the power spectrum there is one high-power region and three low-power regions. High-power regions are preferably assigned to the cell-edge users, while cell-center users are typically served in the low-power regions. As the high-power regions in the neighboring cells are non-overlapping, intercell interference at cell-edge users is received from the low-power regions in the neighboring cells. This improves the cell-edge SINR, while degrading SINR for users at the cell-center. Since a cell-edge user typically experiences a low SINR, its data rate would increase almost linearly with SINR. On the other hand, the degradation in SINR for a high SINR user would only result in a logarithmic reduction of its data rate (see (2.11)). Note that, soft frequency reuse is considered as a universal frequency reuse scheme as it is possible to transmit over all the bandwidth in each cell.

2.3.2 Cell Sectoring

Another well-known method to mitigate the intercell interference is via cell sectoring. This is achieved by using several directional antennas instead of a single omni-directional antenna in each cell. Dividing the cells into sectors actually reduces the network capacity because the available subchannels in

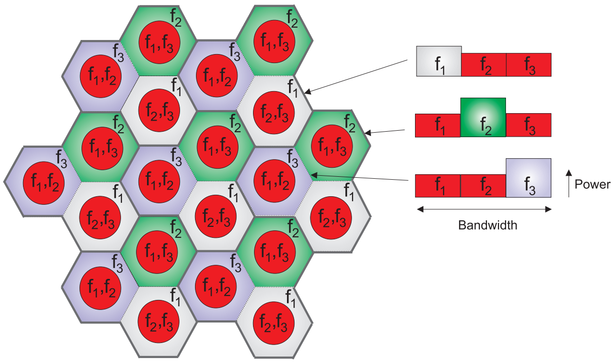


Figure 2.6: An example of soft frequency reuse.

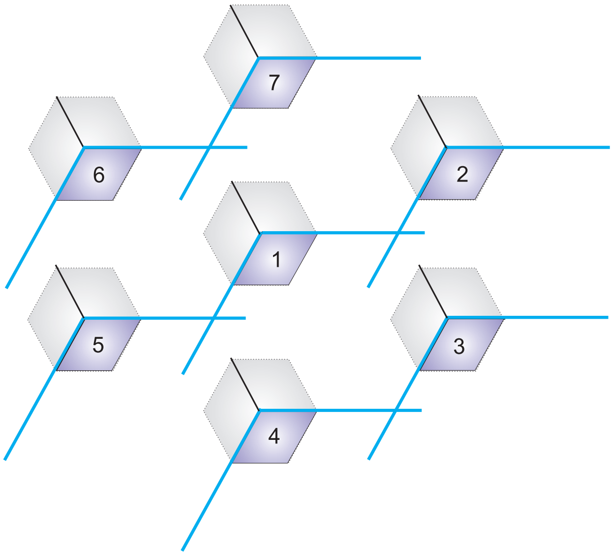


Figure 2.7: An example of cell sectoring with 120° sectors.

each cell are now divided among different sectors (similar to static frequency reuse). The gain in network capacity is, however, achieved by reducing the number of interfering cells. If the same subchannels are assigned to a fixed sector position in all cells, each sector causes interference to the cells that are in its transmission angle only. With no sectoring, intercell interference is received from six neighboring cells in the first tier, while with 120° sectoring, only two neighboring cells interfere with any given sector. This is illustrated in Fig. 2.7, where Sector 1 receives interference only from Sectors 6 and 7.

2.3.3 Cell Densification

Another efficient approach to combat the intercell interference and meet traffic and data rate demands in cellular networks is through increasing the number of cells per unit area. Assuming the number of users in the coverage area remains unchanged, deploying a larger number of base stations will increase the system capacity. This happens because of two main reasons: 1) with more base stations, each user is more likely to find an idle base station to connect to. In other words, each base station has to share its resources with fewer users; 2) as the coverage area of each cell is reduced, the transmit power from base stations can be lowered, resulting in a reduced intercell interference. Denser base station deployment, however, comes at a cost of installing extra towers and backhaul connections as well as extra signaling to support for handoff. Hence, cell densification is sometimes referred to as a *hardware* approach to increase the system capacity. A *software* solution is implemented via multicell coordination, in which the neighboring base stations jointly coordinate the transmission to mitigate or even exploit the intercell interference [13]. This technique is known as base station coordination and will be explained in Chapter 4. Before that, we provide a brief introduction to the use of multiple antennas in conventional non-coordinated networks in Chapter 3.

Chapter 3

Multi-Antenna Cellular Networks

The traditional way of increasing the data rate in cellular networks with a single antenna at each end of the communication link, a.k.a. single-input single-output (SISO), is to increase either the power or the bandwidth or both. However, there are some constraints on the amount that the transmit power and the bandwidth can be increased, which limits the achievable spectral efficiency in SISO networks. From one hand, the data rate of SISO wireless networks can not be increased unboundedly by just increasing the transmit power as these networks are interference-limited. On the other hand, this data rate can not be increased by using a larger bandwidth either, since the spectrum is a limited and expensive resource. The idea of using multiple antennas at both the transmitter and the receiver, referred to as MIMO, serves as an alternative way of increasing the data rate without the need for extra power or bandwidth. The key feature of MIMO techniques is to exploit the degrees of freedom provided by the multipath fading channel, which deteriorates the quality of transmission in traditional SISO networks, to increase the spectral efficiency, suppress interference, and improve the reliability of wireless transmission. In this chapter, a brief overview of different methods to deploy multiple antennas in a cellular network is presented and the performance gain of each method is discussed.

3.1 Single-User MIMO

In a single-user MIMO (SU-MIMO) link, a.k.a. point-to-point MIMO link, one transmitter equipped with multiple antennas communicates with one multi-antenna receiver, as illustrated in Fig. 3.1. In the pioneering work by [14,15,16], it was shown that the use of N_t transmit antennas and N_r re-

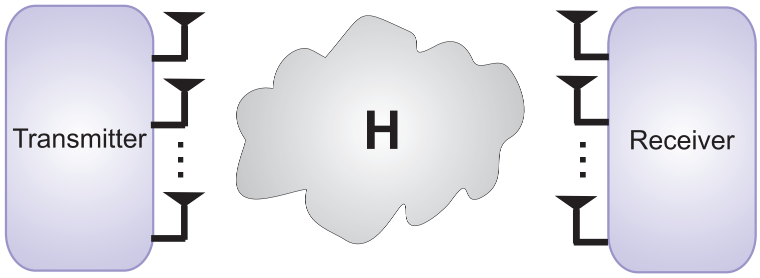


Figure 3.1: Illustration of a point-to-point MIMO link.

ceive antennas can result in a capacity that scales linearly with $\min(N_\mathsf{t}, N_\mathsf{r})$ relative to the SISO case. This capacity increase is achieved under sufficiently rich scattering environment, so that independent transmission paths can be created from each transmit antenna to each receive antenna [17].

3.1.1 MIMO Leverages

In this subsection, we discuss the leverages of SU-MIMO transmission in two different signal-to-noise ratio (SNR) regimes, namely *low SNR* and *high SNR*.

- **Low SNR Regime:** The use of multiple antennas provides an opportunity for beamforming in order to improve the SNR at the receive side. An increase to the average SNR, called *array gain*, as well as a reduction in the SNR fluctuations, known as *diversity gain*, can be made through beamforming at the transmitter or the receiver or both. Beamforming at the transmitter (receiver) requires the CSI to be available at the transmitter (receiver). However, diversity gain at the transmitter can be achieved in the absence of CSI at the transmitter by using a technique known as space-time coding [18, 19]. In contrast to beamforming, space-time coding just reduces the variations in the received SNR and does not improve the average SNR.

- **High SNR Regime:** The use of multiple antenna techniques to improve the received SNR, allows for an increase in the achievable data rates only as long as the data rates are power-limited rather than bandwidth-limited. Therefore, at high enough SNR levels the achievable data rates start to saturate, giving diminishing returns for further increase in the SNR unless the bandwidth is also increased. This can be better understood by considering the basic expression of channel capacity[1] [10]

$$C = \log_2 (1 + \mathsf{SNR}) \,. \tag{3.1}$$

It can be observed from (3.1) that for large values of SNR, the capacity grows logarithmically with the SNR. However, in the case of multiple antennas at both sides of the communication link, it is possible to create multiple parallel data pipes (under certain condition) to share the SNR and avoid the saturation problem. The capacity will grow essentially linearly with the minimum of the number of transmit and receive antennas, i.e., we have $C = \min(N_\mathsf{t}, N_\mathsf{r}) \log_2(\mathsf{SNR}) + O(1)$. This linear increase in capacity resulting from employing multiple antennas at both ends of the communication link is usually referred to as *spatial multiplexing gain* [17].

3.1.2 Capacity of MIMO Channels

In this subsection we investigate the capacity of single-user MIMO channels. We begin by introducing the system model.

System Model

The flat-fading channel matrix between a transmitter with N_t antennas and a receiver with N_r receive antennas is represented by a matrix $\mathbf{H} \in \mathbb{C}^{N_\mathsf{r} \times N_\mathsf{t}}$. The received signal $\mathbf{y} \in \mathbb{C}^{N_\mathsf{r} \times 1}$ can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.2}$$

where $\mathbf{x} \in \mathbb{C}^{N_\mathsf{t} \times 1}$ is the transmitted signal vector, and $\mathbf{n} \in \mathbb{C}^{N_\mathsf{r} \times 1}$ is the additive white Gaussian noise (AWGN) vector with the elements that are distributed as $\mathcal{CN}(0, \sigma_n^2)$. We consider a total average transmitted power constraint as $\mathbb{E}[\mathsf{tr}\{\mathbf{x}\mathbf{x}^\mathsf{H}\}] \leq P$, where $\mathbb{E}[\cdot]$ denotes the expectation operator and $\mathsf{tr}\{\cdot\}$ is the trace operator.

Capacity of Constant MIMO Channels

The capacity of a single-user time-invariant channel is defined by Shannon's capacity theorem to be the maximum data rate that can be transmitted over the channel with arbitrarily small error probability [17]. Channel capacity is defined with no limits on computational complexity or delay, and thus is a fundamental measure of the performance limits of any communication system. It was shown in [16] that if perfect CSI is available both at the transmit and the receive sides, the MIMO channel can be transformed into

[1] For notational convenience, we drop the time-slot index t temporarily until Section 3.2.2.

parallel sub-channels by using singular value decomposition (SVD) of the channel. The capacity of the constant MIMO channel is given by

$$C = \max_{\mathbf{Q}:\mathsf{tr}\{\mathbf{Q}\}\leq P} \log_2 \det\left(\mathbf{I}_{N_\mathsf{r}} + \frac{1}{\sigma_n^2}\mathbf{H}\mathbf{Q}\mathbf{H}^\mathsf{H}\right), \tag{3.3}$$

where $\mathbf{Q}$ is the covariance matrix of the input signal, i.e., $\mathbf{Q} = \mathbb{E}[\mathbf{x}\mathbf{x}^\mathsf{H}]$. Assuming the CSI is available at the transmitter, we can use the eigen value decomposition to diagonalize the Hermitian matrix $\mathbf{H}\mathbf{H}^\mathsf{H}$ as $\mathbf{H}\mathbf{H}^\mathsf{H} = \mathbf{U}^\mathsf{H}\mathbf{\Lambda}\mathbf{U}$. Here, $\mathbf{U}$ is a unitary matrix and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{N_\mathsf{r}})$ is a non-negative diagonal matrix, with λ_i, for $i = 1, 2, \ldots, N_\mathsf{r}$, being the i-th eigenvalue of $\mathbf{H}\mathbf{H}^\mathsf{H}$. Now, using the matrix identity $\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})$ twice, (3.3) can be rewritten as [17]

$$C = \max_{\mathbf{Q}:\mathsf{tr}\{\mathbf{Q}\}\leq P} \log_2 \det\left(\mathbf{I}_{N_\mathsf{r}} + \frac{1}{\sigma_n^2}\mathbf{\Lambda}^{1/2}\mathbf{U}\mathbf{Q}\mathbf{U}^\mathsf{H}\mathbf{\Lambda}^{1/2}\right). \tag{3.4}$$

Note that $\tilde{\mathbf{Q}} = \mathbf{U}\mathbf{Q}\mathbf{U}^\mathsf{H}$ is non-negative definite if and only if $\mathbf{Q}$ is non-negative definite. Furthermore, $\mathsf{tr}(\tilde{\mathbf{Q}}) = \mathsf{tr}(\mathbf{Q})$, which means that the maximization over $\mathbf{Q}$ can be done equally well over $\tilde{\mathbf{Q}}$. We next use the matrix identity $\det(\mathbf{B}) \leq \prod_i \mathbf{B}_{ii}$, for every non-negative definite matrix $\mathbf{B}$, to write

$$\det\left(\mathbf{I}_{N_\mathsf{r}} + \frac{1}{\sigma_n^2}\mathbf{\Lambda}^{1/2}\tilde{\mathbf{Q}}\mathbf{\Lambda}^{1/2}\right) \leq \prod_i (1 + p_i\lambda_i), \tag{3.5}$$

where p_i denotes the i-th diagonal element of $\tilde{\mathbf{Q}}$. Notice that the equality in (3.5) occurs when $\tilde{\mathbf{Q}}$ is diagonal. In this case, the optimal power allocation matrix is found through waterfilling as

$$p_i = \left(\mu - \frac{\sigma_n^2}{\lambda_i}\right)^+, \tag{3.6}$$

where x^+ is defined as $\max(x, 0)$, and μ is the waterfill level chosen such that the total power constraint is satisfied, i.e.,

$$\sum_{i=1}^{N_\mathsf{r}} \left(\mu - \frac{\sigma_n^2}{\lambda_i}\right)^+ = P. \tag{3.7}$$

Acquiring CSI at the transmitter is not easy in general. Assuming no CSI at the transmitter, a natural choice for the input covariance matrix would be $\mathbf{Q} = (P/N_\mathsf{t})\mathbf{I}_{N_\mathsf{t}}$, meaning that the signals transmitted from different antennas are independent and have equal power.

Capacity of Time-Varying MIMO Channels

In the case of a time-varying channel, definition of the capacity depends on the availability of CSI at either the transmitter or the receiver or both. In low mobility scenarios, it can be assumed that the fading is slow enough such that the channel can be estimated reliably at the receiver and fed back to the transmitter with negligible delay. In moderate to high mobility scenarios, however, the channel changes might not be tractable anymore. A common model is to assume a *block-fading* channel model [20]. According to this model, the channel coefficients are constant over a so-called fading coherence block with length of Δ complex dimensions (or channel uses) and changes independently from one block to another. The fading coherence block length Δ in the time-frequency domain is proportional to the product $W_\mathsf{c}T_\mathsf{c}$, where T_c, measured in seconds, denotes the channel coherence time, and W_c, measured in Hertz, is the channel coherence bandwidth [21]. In particular, the high-SNR capacity of a single-user MIMO system with N_t transmit antennas and N_r receive antennas over a block-fading channel with coherence block of length Δ complex dimensions, scales as $C = \min\{N_\mathsf{t}, N_\mathsf{r}, \Delta/2\} \log(SNR) + O(1)$ [22, 23]. Therefore, even when the number of transmit and receive antennas grows very large, i.e., $N_\mathsf{t}, N_\mathsf{r} \gg 1$, the available degrees of freedom are eventually limited by the fading coherence block length Δ. Note that, the coherence block length is usually assumed to be large enough such that very powerful capacity-achieving codes can be used. Depending on the length of the transmitted codeword two different capacity measures can be defined as follows.

- **Ergodic Capacity:** If the transmitted codeword span an infinite number of fading blocks, then the ergodic capacity is a suitable performance metric. In this scenario, assuming the availability of perfect CSI at both the transmitter and the receiver, the ergodic capacity is defined as the average of the capacity achieved for each realization of the channel explained in the previous section, i.e.,

$$C = \mathbb{E}_{\mathbf{H}} \left[\max_{\mathbf{Q}:\mathsf{tr}\{\mathbf{Q}\} \leq P} \log_2 \det(\mathbf{I}_{N_\mathsf{r}} + \frac{1}{\sigma_n^2} \mathbf{H}\mathbf{Q}\mathbf{H}^\mathsf{H}) \right]. \tag{3.8}$$

  On the other hand, if $\mathbf{H}$ contains i.i.d. elements and is only known to the receiver but not to the transmitter, it is shown in [24, 16] that the optimum input covariance matrix that maximizes the ergodic capacity is the identity matrix scaled by the transmit power divided equally among all transmit antennas.

- **Outage Capacity:** If the transmitted codeword spans a single fading block and CSI is perfectly available at the receiver, but not at the transmitter, outage capacity is a relevant measure. In this situation the Shannon capacity is zero due to the fact that no matter how small the rate at which the communication is performed, there is always a nonzero probability that the given channel realization will not support this rate [9]. The $b\%$ outage capacity C_{out}^{b}, is defined as the rate that is guaranteed for $(100-b)\%$ of the channel realizations [20, 25], i.e.,

$$\mathsf{Pr}(C < C_{\mathsf{out}}^{b}) = b\%. \tag{3.9}$$

3.2 Single-Cell Multiuser MIMO Communications

Although SU-MIMO techniques have shown to achieve a capacity that scales linearly with $\min(N_{\mathsf{t}}, N_{\mathsf{r}})$, this gain is based on some premise such as rich-scattering propagation environment. Furthermore, in a cellular system, it is possible to deploy a large number of antennas at the base station, while user terminals usually have a smaller number of antennas due to constraints on size and cost. Therefore, in transmission schemes such as time division multiple access (TDMA), where only one user communicates at each time-slot with the base station, the multiplexing gain $\min(N_{\mathsf{t}}, N_{\mathsf{r}}) = N_{\mathsf{r}}$ is limited by the number of antennas at the user terminals. An alternative to SU-MIMO is the multiuser MIMO (MU-MIMO). In this technique, the spatial degrees of freedom provided by multiple antennas at the base station can be exploited to simultaneously serve multiple users, each with single or multiple antennas, over the same time-frequency resource. Several key advantages of MU-MIMO over SU-MIMO are as follows:

- In MU-MIMO the multiplexing gain[2] is achieved without the need for multi-antenna terminals, thereby allowing the development of small and cheap terminals. This allows to keep the intelligence and cost on the infrastructure side.

- MU-MIMO is more immune to the limitations imposed by the propagation environment such as poor scattering. As a matter of fact, line-of-sight condition which completely degrades the multiplexing gain in SU-MIMO is no longer a problem in MU-MIMO.

In MU-MIMO transmission, a new form of interference resulting from simultaneous transmission to/from multiple users over the same time-frequency slot emerges. This form of interference is commonly referred to as inter-user interference (or intracell interference in cellular context). As in MU-MIMO communication no coordination is assumed among users, inter-user interference mitigation is different in uplink and downlink.

[2]Note that the multiplexing gain in MU-MIMO is sometimes called *multiuser multiplexing gain.*

In the uplink transmission, there is little to be optimized at the user side due to the lack of coordination among the users. Therefore, the challenge is for the base station to separate the signals transmitted by different users. This can be easily attained assuming the base station has perfect CSI knowledge of all users and implements a classical multiuser receiver to separate the signals of different users.

In the downlink transmission, on the other hand, inter-user interference mitigation at the user side is too complicated due to limitations such as size and power at user terminals (which do not exist at the base station). Hence, in the downlink, it is preferred to move the inter-user interference mitigation to the base station side [26]. This is a challenging task since it requires the knowledge of CSI of all users at the base station to properly serve spatially multiplexed users. Note that in the extreme case of no CSI at the base station and identical fading statistics at all receivers, the multiuser multiplexing gain will be lost [27]. Therefore, although not necessary for SU-MIMO transmission, CSI has a cardinal role in inter-user interference mitigation techniques for MU-MIMO transmission. Hence, we only focus on the downlink in the rest of the thesis, where base stations are the transmitters and users are the receivers.

We consider a base station equipped with N_t antennas communicating with $K \geq N_\mathsf{t}$ users, each with N_r antennas. The received signal at the k-th user in such a system can be expressed as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k, \tag{3.10}$$

where $\mathbf{y}_k \in \mathbb{C}^{N_\mathsf{r} \times 1}$ is the received signal vector of user k, $\mathbf{H}_k \in \mathbb{C}^{N_\mathsf{r} \times N_\mathsf{t}}$ denotes the downlink channel matrix between the base station and user k, $\mathbf{x} \in \mathbb{C}^{N_\mathsf{t} \times 1}$ is the transmit signal vector, and $\mathbf{n}_k \in \mathbb{C}^{N_\mathsf{r} \times 1}$ denotes the AWGN vector at user k. The transmit signal vector $\mathbf{x}$ is in general a function of each user k's data signal vector $\mathbf{x}_k$, for $k = 1, \ldots, K$. The covariance matrix of user k transmit signal vector is given by $\mathbf{Q}_k = \mathbb{E}[\mathbf{x}_k \mathbf{x}_k^\mathsf{H}]$. Furthermore, we assume a total power constraint of P at the base station, i.e., $\sum_{k=1}^{K} \mathsf{tr}(\mathbf{Q}_k) \leq P$.

In a SU-MIMO system, the capacity of the system is characterized by one single number. In a MU-MIMO system with K users, however, the capacity of the system is characterized by a region of K dimensions, denoted as the *capacity region*, in which each K-dimensional vector represents the achievable rates by all K users. A rate vector is said to be achievable if there exists a coding scheme for which the joint probability of error for all users can be made arbitrarily small as the code block length becomes sufficiently large. Hence, in a MU-MIMO system with homogenous users, i.e., users with the same average SNR, one common performance metric to optimize is the *sum rate* of all users.

It has been shown that the optimal transmission strategy for a single base station to serve multiple users simultaneously (a.k.a. MIMO broadcast channel) is through a coding scheme known as dirty paper coding (DPC) [28]. This coding scheme is based on interference pre-subtraction combined with an implicit user scheduling and power loading algorithm. In particular, it can be shown that when the number of users K is larger than the number of base station antennas N_{t}, capacity scales linearly with N_t using DPC. The concept of DPC was first introduced by Costa [29]. He showed that for a scalar AWGN channel with an interfering Gaussian signal known non-causally at the transmitter (but not at the receiver), the capacity is the same as if there was no additive interference, or equivalently as if the receiver also had the knowledge of interference. In other words, if the interference is known non-causally, it can be pre-subtracted at the transmitter with no increase in the transmit power. This idea was extended to the MIMO case later and in [28] it was shown that DPC is in fact the capacity-achieving transmission strategy in the MIMO broadcast channel. Although theoretically optimal, DPC is very difficult to implement in practical systems due to high computational burden of successive encodings and decodings, especially when the number of users K is large.

To operate at a point on the capacity region boundary, the base station needs to serve all K users simultaneously. Note that the allocated resources to each user, e.g., power, depends on the instantaneous channel conditions and may vary greatly from user to user. Furthermore, the multiplexing gain is limited by the number of antennas at the base station. Hence the number of users that are effectively served with non-zero power at any given instant of time is directly related to N_{t}. This number can be considerably less than the total number of users K. Studies have shown that the maximum number of users with non-zero allocated power at any given realization of the channel is upper bounded by N_{t}^2 [30]. Users with non-zero allocated power at each time-slot are usually referred to as *active*, *scheduled*, or *selected* users. The complexity of DPC implementation from one side and the issue of the optimal number of users to serve on the other side, motivate the need for more practical transmission strategies and user scheduling algorithms.

3.2.1 Linear Precoding

The operation of precoding at the transmitter side is similar to that of equalization at the receiver side. As a result, precoding requires an accurate estimate of CSI at the transmitter. Several low-complexity linear and non-linear precoding schemes have been proposed that when combined with single user encoding and decoding, can achieve near capacity performance. In particular, it has been shown that linear beamforming schemes asymptotically approach DPC performance [31]. For a MU-MIMO system with homogeneous users this means that the sum-rate scales as $N_{\mathsf{t}} \log(\mathsf{SNR})$ for SNR approaching infinity with fixed $K \gg N_{\mathsf{t}}$, and scales as $N_{\mathsf{t}} \log \log K$ for K going to infinity with fixed SNR. Linear precoding, however, might perform far from DPC when the number of users is small. Non-linear precoding techniques such as vector perturbation [32] and Tomlinson-Harashima [33] can perform closer to DPC even when the number of users is small by performing additional signal processing at the transmitter. In this thesis, we assume the number of users is much larger than the number of transmit antennas whenever we consider user scheduling. Moreover, we only focus on linear precoding techniques for simplicity. In the rest of this subsection, we provide an overview of three popular linear precoding techniques, namely zero-forcing (ZF), minimum mean square error (MMSE), and block diagonalization (BD).

In the classical space division multiple access (SDMA) scheme with single-antenna users, as illustrated in Fig. 3.2, each user's stream is encoded independently and multiplied by a so-called beamforming (BF) weight vector for transmission through multiple antennas. This will reduce (or eliminate) the interference among different streams due to spatial separation of users. Linear precoding is a generalization of SDMA, where each user is equipped with multiple antennas and therefore is assigned a precoding matrix instead of a BF vector. For the MU-MIMO system under consideration, define $\mathbf{s}_k \in \mathbb{C}^{N_\mathsf{t} \times 1}$ to be the transmitted signal vector for user k. This signal vector is multiplied by a precoding matrix $\mathbf{T}_k \in \mathbb{C}^{N_\mathsf{t} \times N_\mathsf{r}}$ and sent to the base station antennas for transmission. The transmit signal from N_t antennas at the base station is therefore given by

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{T}_k \mathbf{s}_k. \tag{3.11}$$

The received signal vector in (3.10) can then be rewritten as

$$\mathbf{y}_k = \underbrace{\mathbf{H}_k \mathbf{T}_k \mathbf{s}_k}_{\text{desired signal}} + \underbrace{\sum_{j=1, j\neq k}^{K} \mathbf{H}_k \mathbf{T}_j \mathbf{s}_j}_{\text{inter-user interference}} + \mathbf{n}_k, \tag{3.12}$$

where, in the right hand side of the equality (3.12), the first term is the desired signal and the second term denotes the inter-user interference. The precoding matrices are designed jointly based on a number of optimization criteria and the availability of CSI from all users. Under single-user detection, where each user only decodes its own data signal, while treating inter-user interference as part of the background noise, the achievable rate

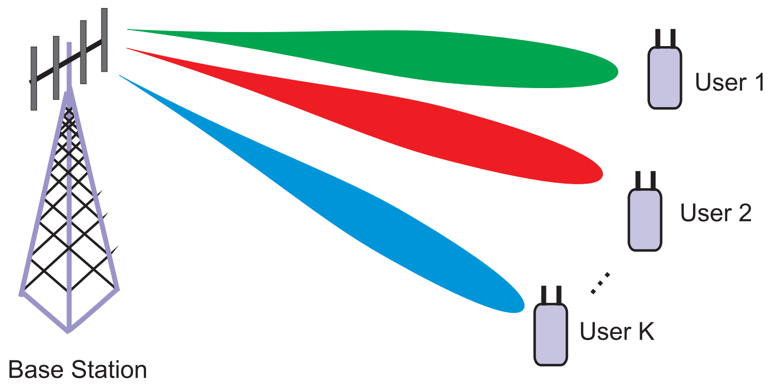


Figure 3.2: Schematic illustration of SDMA.

of user k is given by [34]

$$\begin{aligned} R_k &= I(\mathbf{s}_k; \mathbf{y}_k) \\ &= \log_2 \det \left(\mathbf{I} + \left[\mathbf{I} + \sum_{j=1, j\neq k}^{K} \mathbf{H}_k \mathbf{T}_j \mathbf{Q}_j \mathbf{T}_j^{\mathrm{H}} \mathbf{H}_k^{\mathrm{H}} \right]^{-1} \mathbf{H}_k \mathbf{T}_k \mathbf{Q}_k \mathbf{T}_k^{\mathrm{H}} \mathbf{H}_k^{\mathrm{H}} \right), \end{aligned} \tag{3.13}$$

where $I(\mathbf{s}_k; \mathbf{y}_k)$ denotes the mutual information between $\mathbf{s}_k$ and $\mathbf{y}_k$. An example of an optimization criterion for designing precoding matrices is to maximize the sum rate of all users subject to a total power constraint at the base station. In this case, one should maximize R_{sum}, given by

$$R_{\mathsf{sum}} = \max_{\mathbf{T}_k, \mathbf{Q}_k} \sum_{k=1}^{K} R_k, \tag{3.14}$$

$$\text{subject to: } \sum_{k=1}^{K} \mathsf{tr}(\mathbf{T}_k \mathbf{Q}_k \mathbf{T}_k^{\mathsf{H}}) \leq P. \tag{3.15}$$

Channel Inversion

Channel inversion, or ZF, is a sub-optimal precoding strategy that can be implemented easily in practice with a performance comparable to that of DPC in some cases. In linear precoding the number of users which can be served simultaneously is limited by the number of degrees of freedom, which is equal to N_{t}. Usually a subset $\mathcal{S}$ of users is selected for transmission by the scheduler at each time-slot (This will be explained in more detail in the next subsection). For ease of exposition, we focus on $N_{\mathsf{r}} = 1$ and denote the channel and the BF vector of user k as $\mathbf{h}_k$ and $\mathbf{t}_k$, respectively. To satisfy the dimensionality constraint of linear precoding we should have $|\mathcal{S}| \leq N_{\mathsf{t}}$. Let $\mathbf{H} = [\mathbf{h}_1^{\mathsf{H}} \dots \mathbf{h}_K^{\mathsf{H}}]^{\mathsf{H}}$ denote the aggregate channel matrix, and $\mathbf{T} = [\mathbf{t}_1^{\mathsf{H}} \dots \mathbf{t}_K^{\mathsf{H}}]^{\mathsf{H}}$ be the aggregate precoding matrix of all users. For the scheduled users in $\mathcal{S}$, denote $\mathbf{H}(\mathcal{S})$ and $\mathbf{T}(\mathcal{S})$ as the corresponding submatrices of $\mathbf{H}$ and $\mathbf{T}$, respectively. The ZF precoding is found by finding the Moore-Penrose pseudo inverse of $\mathbf{H}(\mathcal{S})$ as [31]

$$\mathbf{T}(\mathcal{S}) = \mathbf{H}(\mathcal{S})^{\dagger} = \mathbf{H}(\mathcal{S})^{\mathsf{H}} \left(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^{\mathsf{H}}\right)^{-1}. \tag{3.16}$$

The achievable sum rate is given by

$$R_{\mathsf{sum}} = \max_{\substack{p_k \\ \sum_{k \in \mathcal{S}} \eta_k p_k \leq P}} \sum_{k \in \mathcal{S}} \log(1 + p_k), \tag{3.17}$$

where p_k is the allocated power to the k-th user and

$$\eta_k = \frac{1}{\left[\left(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^{\mathsf{H}}\right)^{-1}\right]_{k,k}} \tag{3.18}$$

denotes the effective channel gain for user k. The transmit power allocation depends on the optimization criterion, e.g., for sum rate maximization the optimal power allocation is achieved via a waterfilling algorithm as explained in Section 3.1.1. In general, ZF is not power efficient since the weight vector of each user does not match its channel vectors. In other words, when the channel is rank-deficient, it causes signal attenuation at the base station side. As a result, the capacity of channel inversion does not increase linearly with N_{t}. However, if the number of users, i.e., K, is asymptotically large and user selection is performed, the sum rate of ZF precoding approaches that of DPC [31]. This is due to *multiuser diversity*, which is a form of selection diversity that gives the opportunity to the base station to select a subset of users with favorable channel conditions. There are two major benefits provided by multiuser diversity, namely increased channel magnitudes and abundant channel directions. As an example of the former benefit, consider a set of K homogeneous users. The channel gain of the best user is then roughly $\log K$ times higher than the average channel gains, resulting in the multiuser diversity to increase the SNR by the same factor. Furthermore, the provided abundant channel direction enables the base station to choose a user group with good spatial separations, which minimizes the signal attenuation problem in ZF.

Regularized Channel Inversion- MMSE Precoding

Similar to minimum mean square error (MMSE) equalization, regularized channel inversion adds a regularization vector to the ZF precoder, to reduce the signal attenuation problem, i.e.,

$$\mathbf{T}(\mathcal{S}) = \mathbf{H}(\mathcal{S})^{\dagger} = \mathbf{H}(\mathcal{S})^{\mathsf{H}} \left(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^{\mathsf{H}} + \varepsilon \mathbf{I}\right)^{-1}. \tag{3.19}$$

Here, ε is the regularization factor, which is usually determined heuristically to obtain a good tradeoff between the numerical conditioning of the channel inversion and the amount of residual interference [32]. In contrast to the channel inversion scheme, regularized channel inversion leads to linear capacity growth with the number of transmit antennas. Furthermore, the performance of MMSE is much better at low SNR and converges to that of ZF at high SNR. However, it does not completely eliminate the inter-user interference, thus optimal power allocation can not be performed easily.

Block Diagonalization (BD)

When each user has multiple antennas, inversion of the aggregate channel at the transmitter is sub-optimal, since some level of coordination is allowed among the antennas of each user. Let us re-define the aggregate users' channel and precoding matrices as

$$\mathbf{H} = [\mathbf{H}_1^{\mathsf{H}} \quad \mathbf{H}_2^{\mathsf{H}} \quad \cdots \quad \mathbf{H}_K^{\mathsf{H}}]^{\mathsf{H}}, \tag{3.20}$$

$$\mathbf{T} = [\mathbf{T}_1^{\mathsf{H}} \quad \mathbf{T}_2^{\mathsf{H}} \quad \cdots \quad \mathbf{T}_K^{\mathsf{H}}]^{\mathsf{H}}. \tag{3.21}$$

The optimal design for zero inter-user interference is obtained when $\mathbf{HT}$ is block diagonal [35]. In order to achieve this property, the precoding matrices $\{\mathbf{T}_j\}_{j=1}^{K}$ are chosen such that $\mathbf{H}_k\mathbf{T}_j = \mathbf{0}$, $\forall k \neq j$. Let $\tilde{\mathbf{H}}_k \in \mathbb{C}^{(K-1)N_{\mathsf{r}} \times N_{\mathsf{t}}}$ be defined as

$$\tilde{\mathbf{H}}_k = [\mathbf{H}_1^{\mathsf{H}} \quad \cdots \quad \mathbf{H}_{k-1}^{\mathsf{H}} \quad \mathbf{H}_{k+1}^{\mathsf{H}} \quad \cdots \quad \mathbf{H}_K^{\mathsf{H}}]^{\mathsf{H}}. \tag{3.22}$$

Then the zero inter-user interference is obtained by forcing $\mathbf{T}_k$ to lie in the null space of $\tilde{\mathbf{H}}_k$. It reveals that the necessary condition to accommodate all users under a zero-interference constraint is that the null space of $\tilde{\mathbf{H}}_k$ has a dimension greater than zero, i.e., $\mathrm{rank}(\tilde{\mathbf{H}}_k) < N_{\mathsf{t}}$. Let the SVD of $\tilde{\mathbf{H}}_k$ be

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k \tilde{\mathbf{\Sigma}}_k [\tilde{\mathbf{V}}_k^{(1)} \ \tilde{\mathbf{V}}_k^{(0)}]^{\mathsf{H}}, \tag{3.23}$$

where $\tilde{\mathbf{V}}_k^{(1)}$ consists of the principal $\tilde{\ell}_k = \mathrm{rank}(\tilde{\mathbf{H}}_k)$ right singular vectors, corresponding to non-zero singular values. On the other hand, $\tilde{\mathbf{V}}_k^{(0)}$ contains the last $N_{\mathsf{t}} - \tilde{\ell}_k$ right singular vectors corresponding to zero singular values, and thus forms an orthonormal basis for the null space. Let ℓ_k denote the rank of the effective channel matrix for user k, i.e., $\mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}$. A sufficient condition to make the transmission to user k feasible, i.e., to have $\ell_k \geq 1$, is that at least one row of $\mathbf{H}_k$ is linearly independent of the rows of $\tilde{\mathbf{H}}_k$ [35]. The channel of user k after projection to the null space of $\tilde{\mathbf{H}}_k$ becomes inter-user interference free. Therefore, the effective channel $\mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}$ is similar to a SU-MIMO channel, for which the optimal transmission strategy is the one based on SVD precoding and waterfilling over its singular values [16] (see also Section 3.1.1). Define the SVD of the effective channel for user k as

$$\mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)} = \mathbf{U}_k \begin{bmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{V}_k^{(1)} \ \mathbf{V}_k^{(0)}]^{\mathsf{H}}, \tag{3.24}$$

where $\mathbf{\Sigma}_k$ contains the ℓ_k non-zero singular values and $\mathbf{V}_k^{(1)}$ the corresponding ℓ_k singular vectors. The precoding matrix $\mathbf{T}_k$ can then be chosen as the product of $\tilde{\mathbf{V}}_k^{(0)}$ and $\mathbf{V}_k^{(1)}$.

3.2.2 Downlink Resource Allocation

Resource allocation techniques play an important role in efficient utilization of the scarce radio resources and provision of the quality of service in the downlink of multiuser MIMO networks. In a very general sense, a downlink resource allocation policy is defined by a *physical (PHY) layer signaling* $\mathcal{P}$ and a *MAC layer scheduling algorithm* ω. The PHY layer signaling determines the type of coding used for transmission, such as superposition coding, spatial precoding, etc. The MAC layer scheduling at each time-slot t decides which user must be served and with what rate, by choosing a rate K-tuple $\mathbf{R}(t) = [R_1(t) \cdots R_K(t)]$. This rate vector is a function of CSI and some other fairness criteria (more details will be specified later). Let $\mathcal{R}_{\mathcal{P}}(t)$ denote the instantaneous rate region during time-slot t. For all rate vectors $\mathbf{R}(t) \in \mathcal{R}_{\mathcal{P}}(t)$, communication to all users is performed with vanishing error probability. A *feasible* scheduling algorithm for the signaling scheme $\mathcal{P}$ is one that chooses $\mathbf{R}(t) \in \mathcal{R}_{\mathcal{P}}(t)$ for all t.

Let $\bar{R}_k$ denote the long-term average rate (a.k.a. throughput) of a user k in bits/s/Hz, defined as

$$\bar{R}_k = \eta_k R_k. \tag{3.25}$$

Here, η_k denotes the fraction of time-slots on which user k is served, referred to as the *activity fraction*, and R_k is the average rate of user k if it was served at all time-slots, nicknamed as *net average rate* hereafter. The value of η_k is determined by the scheduling algorithm in the MAC layer, while the value of R_k depends on the SINR experienced by the user. The achievable *ergodic* rate region (or the throughput region) for a given signaling scheme $\mathcal{P}$ is defined as

$$\mathcal{R}_{\mathcal{P}} = \text{coh} \bigcup_{\omega \in \Omega_{\mathcal{P}}} \left\{ \bar{\mathbf{R}} \in \mathbb{R}_{+}^{K} : \bar{R}_k \leq \liminf_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} R_k(t), \forall k \right\}, \tag{3.26}$$

where "coh" denotes the "closure of the convex hull" and $\Omega_{\mathcal{P}}$ is the set of all feasible scheduling policies under the signaling scheme $\mathcal{P}$. The goal of downlink scheduling is to make the system operate at a desired point on the ergodic rate region for a given PHY layer signaling scheme. This is achieved by solving the following optimization problem

$$\begin{aligned} &\max U(\bar{\mathbf{R}}) \\ &\text{subject to: } \bar{\mathbf{R}} \in \mathcal{R}_{\mathcal{P}}, \end{aligned} \tag{3.27}$$

where $U(\cdot)$ is a continuous, strictly concave, and component-wise increasing utility function, reflecting some suitable notion of fairness [36].

In high-rate data-oriented downlink systems, two specific settings are of particular interest: 1) systems with random data arrivals and transmission queues; 2) systems with infinite backlog. In the former, the scheduling algorithm should adapt to both the channel and the data traffic variations. The main goal in this case is to achieve stability of the transmission queues, such that they all have a finite average buffer size [37]. In the latter, however, users' data are already available at the transmitter and only the channel is varying. The main objective here is to share the channel among the users to maximize the average throughput subject to some fairness constraint. Note that in a system with homogeneous users, scheduling algorithms that maximize the sum rate (i.e., pick up the users with the best channel conditions) at each time-slot turns out to also maximizes the throughput of individual users. When realistic distance-dependent pathloss and shadowing are taken into account, the average SNR of users are not necessarily equal. In such a scenario, the scheduling strategy that maximizes the sum rate might result in serving the users close to the base station in most of the time-slots, leaving the users with a larger pathloss unserved for a long time. This makes the notion of fairness particularly important in scenarios including realistic distant-dependent pathloss. A simple scheduling algorithm, which has been used in 1xEV-DO and the data-oriented downlink schemes of CDMA2000 [38], to address the fairness issue is the *proportional fairness scheduling* (PFS). In this approach, the scheduling optimization problem is formulated as

$$\begin{aligned} &\max \sum_{k=1}^{K} \log(\bar{R}_k) \\ &\text{subject to: } \bar{\mathbf{R}} \in \mathcal{R}_{\mathcal{P}}. \end{aligned} \tag{3.28}$$

It is well known that the solution to (3.28) is achieved by a dynamic policy that at each time-slot t selects a rate vector $\mathbf{R}^*(t)$ such that

$$\mathbf{R}^*(t) = \arg \max_{\mathbf{R}(t) \in \mathcal{R}_{\mathcal{P}}(t)} \sum_{k=1}^{K} \frac{R_k(t)}{\bar{R}_k}, \tag{3.29}$$

where $\mathcal{R}_{\mathcal{P}}(t)$ denotes the instantaneous achievable rate region for PHY layer signaling $\mathcal{P}$. Note that the ergodic rates $\bar{R}_k$, for $k = 1, \ldots, K$, are generally not known a priori. Hence, an adaptive method is used to estimate these values recursively as [38]

$$\bar{R}_k(t) = (1 - 1/T)\bar{R}_k(t-1) + (1/T)R_k(t), \tag{3.30}$$

where $\bar{R}_k(t)$ denotes the average rate until time-slot t, and $1/T$ is a constant that governs the size of a exponential moving average window. Note that $\bar{R}_k(t-1)$ is used in (3.29) in place of $\bar{R}_k$ at each time-slot t. Furthermore, it is known that the adaptive algorithm approaches the exact PFS rates in the limit of very small $1/T$.

Under PHY layer signaling based on linear precoding and assuming single-antenna users, the optimization problem in (3.29) is solved via finding a subset of users that maximize the weighted sum rate. This can be done with a brute-force complete search over all possible combinations of users. When the number of users is large, however, the computational complexity of the brute-force approach becomes prohibitive. As a result, low-complexity greedy user selection schemes have been proposed in [39, 31, 40]. For example, in the capacity-based greedy user selection algorithm, the transmitter chooses the single user with the highest rate. Then, it finds the next user among the remaining unselected users that provides the maximum sum rate (or weighted sum rate) with the previously selected users. The algorithm is repeated until either there is no more improvement in the sum rate or the number of selected users is equal to the number of transmit antennas.

To clarify the operation of PFS for homogeneous and heterogeneous users, we consider two cases where a base station with 2 antennas serves 4 single-antenna users using ZF beamforming. In the first case, all users have an equal average SNR of 15 dB. In the second scenario, all 4 users are placed over a line at distances of 100, 400, 700, and 1000 m from the base station, respectively. The pathloss model for user k is given by $(d_k/D_0)^{-\upsilon}$, where d_k is the distance of user k to the base station, $D_0 = 1000$ m denotes a reference distance, and $\upsilon = 3.5$ is the pathloss exponent. Furthermore, the transmit power at the base station in this scenario is chosen such that the farthest user experiences an average SNR of 15 dB. Figure 3.3 shows the user throughput estimates $\bar{R}_k(t)$ in (3.30) versus the time-slot index for both the homogeneous (Fig. 3.3(a)) and the heterogeneous (Fig. 3.3(b)) cases. It is observed that the limiting throughput for all homogeneous users is the same. This means that PFS in this case is reduced to sum rate maximization. For heterogeneous users, however, the limiting values are different depending on the location of the users.

3.2.3 CSI Acquisition at the Base Station

Resource allocation for MU-MIMO systems requires the CSI of all users to be available at the base station. In practice, CSI is usually obtained through some form of *training*, *estimation*, and *feedback*, and is usually imperfect. In time-division duplexing (TDD) systems, uplink and downlink transmissions are performed over the same channel coherence bandwidth. In such systems, CSI can be acquired in the so-called *open-loop mode* via training pilots transmitted by users and employing the reciprocity of the uplink and downlink channel. In frequency-division duplexing (FDD) systems, however, uplink and downlink transmissions take place in different widely separated frequency bands. In this scenario, CSI at the base station is obtained in a so-called *closed-loop mode*, where each user estimates its CSI using the training pilots transmitted over the downlink channel, and then feeds this information back to the base station either in quantized (digital) or unquantized (analog) form over the uplink feedback channel [41,42]. The quality of the CSI at the base station depends on the amount of signal dimensions dedicated to training and feedback. It follows that there is an inherent tradeoff between the advantages of improving the CSI quality and the amount of signal dimensions allocated for CSI estimation and feedback.

In this thesis, we focus on FDD systems, where we assume that there exists error- and delay-free feedback links that convey the user CSI to the base station. Note that in practice, feedback links are subject to both delay and error, which can degrade the performance severely. Analysis of the effect of feedback imperfections are, however, beyond the scope of this thesis. The interested readers are referred to recent works in [43,44,45,46,47,48,49] and the references therein.

3.3 MU-MIMO in Multicell Environment

The predicted gains of MU-MIMO is significantly deteriorated when employed in a multicell environment due to the presence of intercell interference. Note that a multicell MU-MIMO network can be looked at as a conventional cellular network with increased antenna density. The number of antennas in such a system can potentially be equal to that in a cell densified network explained in Section 2.3.3, but with the extra antennas installed in the existing towers instead of new towers (see Fig. 3.4). Assuming the number of users in a given area to remain unchanged, the activity fraction of each user in a multicell MU-MIMO network is increased due to

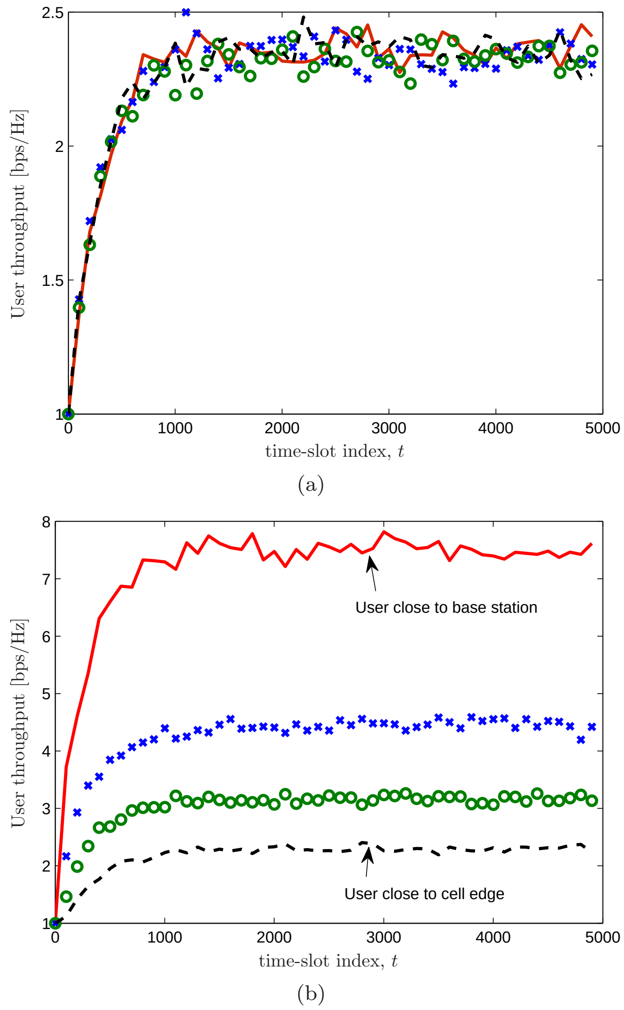
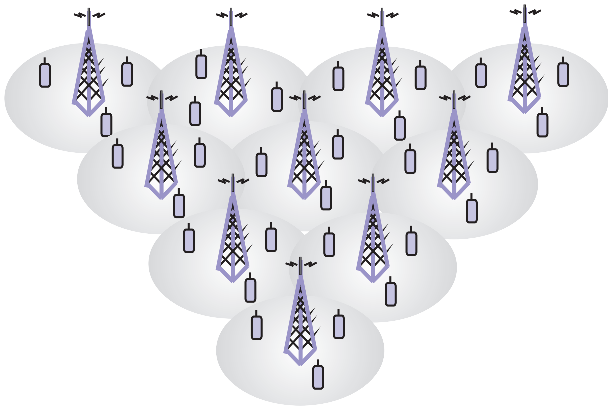


Figure 3.3: User throughput estimates vs. time-slot index for homogeneous (a) and heterogeneous (b) users.

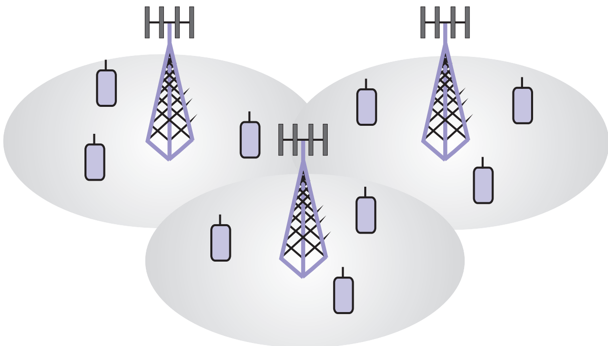
the increased antenna density. The SINR distribution for users is, however, different in such a network compared to that in a cell densified network. Although it is possible to fully coordinate the antennas at each base station in multicell MU-MIMO networks, the users in such networks suffer from intracell interference, which does not exist in cell densified networks. Mitigation of intracell interference at the base station using the techniques discussed in the Section 3.2.1 consumes degrees of freedom, thereby reducing the desired signal level for each user. Furthermore, as the coverage area of each cell is not changed in multicell MU-MIMO networks, the transmit power from base stations can not be lowered as in cell densified networks. Therefore, assuming the same transmit power as in conventional single-antenna networks, the intercell interference power is not reduced in this scenario. However, due to the increased number of interfering signals, the statistical distribution of intercell interference is changed.

We highlight that intercell interference has a fundamental difference from intracell interference caused by other users in the same cell or from interstream interference caused by spatial multiplexing of data streams of a single user. In a network where cells operate independently and concurrently, intercell interference is a non-causal quantity which can not be known prior to transmission. The scheduler at each base station performs user scheduling based only on the CSI knowledge of the users in its own cell (denoted as the local CSI), and without any knowledge about the intercell interference. The intercell interference power observed at each user changes from time-slot to time-slot in a random and unpredictable manner, depending on the scheduling decision made by neighboring cells. Therefore, the instantaneous SINR at each user in the system is a random variable (see eq. (2.10)). Even if MU-MIMO techniques minimize the inter-user interference power using the local CSI, intercell interference power could be large for cell-edge users, resulting in a low SINR values at these users. Hence, although the use of multiple antennas at the base stations can increase the activity fraction of cell-edge users, their performance could still be very poor due to low SINR values. This problem calls for efficient intercell interference mitigation techniques.

In a system with N_I interfering base stations, each with N_t antennas, each user will need $N_\mathsf{r} \geq (N_\mathsf{I}+1)N_\mathsf{t}$ antennas to suppress the spatial interference from the neighboring base stations and decode parallel data streams from its serving base station using linear processing. This is almost impossible to implement in the small size unit at the user side [50]. Therefore, the intercell interference mitigation must be performed at the transmitter side, which does not have such user side constraints on size and power. One of the candidate techniques for this purpose is to exploit or mitigate the intercell interference via coordination among base stations, which is explained in the next chapter.



(a)



(b)

Figure 3.4: An example of a cell densified network with single-antenna base stations (a) and a conventional cellular network with multi-antenna base stations (b).

Chapter 4

Multicell Coordination

As mentioned in the previous section, the performance of MU-MIMO systems in a multicell environment is degraded by intercell interference, especially for users close to the cell edge. One of the efficient techniques to combat the deteriorating effect of intercell interference is to exploit the coordination among base stations. The basic idea behind coordination techniques is to coordinate the transmission to the users in the adjacent cells among their corresponding base stations. This requires the CSI and the user data to be exchanged among these base stations over the backhaul links. Depending on the amount of signaling overhead that can be placed over the backhaul links and the feedback channels, different levels of coordination are defined. Note that the idea of base station coordination is not totally new. In fact, one form of base station coordination already exists in today's 3G network, denoted as soft handoff [51]. In conventional code division multiple access (CDMA) 3G networks, soft-handoff allows a user to communicate simultaneously with several base stations. In this scenario, selection diversity is used to select the best of these connections at any given time. Such selection diversity when combined with power control allows full frequency re-use in each cell. However, full frequency re-use in CDMA network results in per-cell capacity to be constrained by intercell interference [52]. Base station coordination techniques in LTE-Advanced are an attempt to address this intercell interference penalty. In this chapter, an overview of different coordination strategies are presented, and different challenges in implementing each strategy are discussed.

4.1 System Model

A typical coordinated multicell MU-MIMO network comprises B coordinating base stations each with N_{t} antennas. There are K users in each

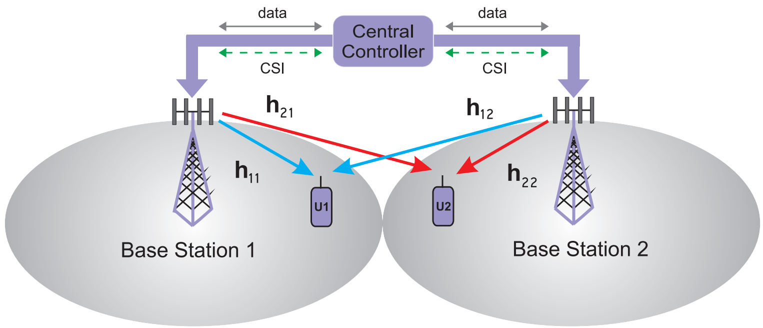


Figure 4.1: A cluster of 2 BSs, which are cooperatively serving 2 users.

cell. Each user can have multiple antennas to enable spatial multiplexing of multiple data streams or intercell interference cancelation at the user side. However, we only consider single-antenna users as we focus on base station side intercell interference mitigation. A narrowband block-fading channel model and universal frequency reuse are considered. The base stations are connected to a central controller via high-speed backhaul links to share data or CSI or both. Figure 4.1 shows an example of a network with 2 coordinating base stations. The complex baseband received signal at user k in cell b, denoted as user k_b, is given by

$$y_{k_b} = \sum_{\ell=1}^{B} \mathbf{h}_{k_b,\ell}^{\mathsf{H}} \mathbf{x}_\ell + z_{k_b}, \tag{4.1}$$

where y_{k_b} is the received signal, $\mathbf{h}_{k_b,\ell} \in \mathbb{C}^{N_\mathrm{t}\times 1}$ denotes the channel vector between user k_b and the base station ℓ, $\mathbf{x}_\ell \in \mathbb{C}^{N_\mathrm{t}\times 1}$ is the transmitted signal from base station ℓ, and z_{k_b} accounts for receiver noise plus any other form of interference coming from outside the B coordinating cells.

4.2 Multicell Coordination Strategies

Depending on the amount of information sharing over the backhaul, multicell coordination techniques can be classified into two main categories, namely *joint processing* and *coordinated scheduling/beamforming* [53].

4.2.1 Joint Processing

In joint processing, user data is available at all the coordinating base stations. One subclass of joint processing is the so-called *network MIMO*, in

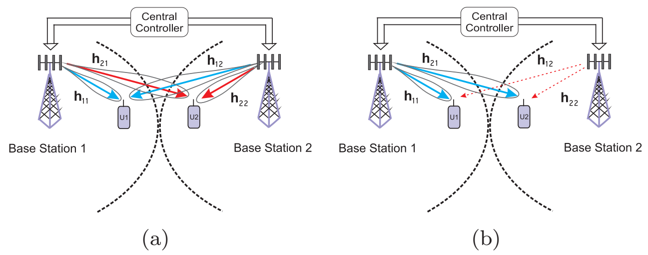


Figure 4.2: Schematic illutration of joint Processing operation: a) network MIMO; b) dynamic cell selection.

which the data to each user is jointly encoded and simultaneously transmitted from all coordinating base stations. Network MIMO operation relies on the assumption that all the coordinating base stations are inter-connected via high-capacity backhaul links and can perfectly share the data and CSI of all users. In this scenario, the concept of serving base station for each user disappears and all the coordinating base stations act as a single distributed multi-antenna transmitter with a per base station power constraint [54]. Therefore, in network MIMO intercell interference is *exploited* as interfering channels are used to transmit useful data. Another subclass of joint processing is *dynamic cell selection* in which the data to each user is transmitted from the coordinating cell with the best channel condition, while the other base stations are muted, so that intercell interference is *mitigated*. The operating principle of network MIMO and dynamic cell selection is shown in Fig. 4.2(a) and Fig. 4.2(b), respectively.

4.2.2 Coordinated Scheduling/Beamforming

In this coordination strategy, user data is only available at one base station, i.e., the *serving* one. It is, however, possible to share the CSI of users via backhaul links to enable the base stations to coordinate their signaling strategies, such as power allocation, beamforming, and user scheduling. No data sharing over the backhaul is required. Note that in this coordination mode, intercell interference is only *mitigated*. The operating principle of coordinated beamforming is shown in Fig. 4.3.

4.3 The Cardinal Role of Scheduling in Multicell Coordination

Scheduling is one of the most important issues in cellular design. As can be seen from (3.25) in Section 3.2.2, the long-term average rate of a user

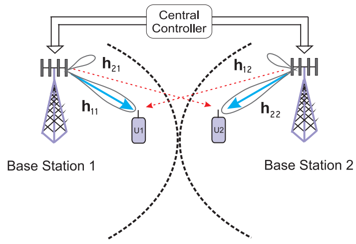


Figure 4.3: Schematic illutration of coordinated beamforming operation.

k in a given cell depends on both its activity fraction, i.e., η_k, and its net average rate, i.e., R_k. The former is usually determined by the scheduler, while the latter is a function of SINR. As a result, focusing on PHY layer techniques such as beamforming design to enhance the SINR does not necessarily improve the achievable throughput. In fact, improving SINR (and hence R_k) might reduce η_k depending on the scheduling criterion. Therefore, in cellular system design, as long as the throughput is the performance metric, it is crucial to consider both the scheduling to control η_k and the SINR enhancement techniques to improve R_k. We clarify the importance of this issue by providing the following example.

EXAMPLE 1: Consider a network consisting of two one-sided linear cells, each with a base station with 2 antennas. Base stations are placed at positions -1 km and 1 km. There are 4 users in each cell equally spaced over the cell area as shown in Fig. 4.4. The user close to the cell-edge in each cell is classified as the cell-edge user (see Fig. 4.4). Users are indexed such that the user 1 is the closest to the base station and user 4 is the closest to the cell-edge. Each base station has the perfect CSI of the users in its own cell and employs PFS together with ZF beamforming to serve at most 2 out of 4 users in its own cell at each time-slot. Let $\bar{b} = \mod(b, 2) + 1$, for $b = 1, 2$, denote the other base station/cell depending on the context and $\mathcal{S}_b(t)$ be the set of selected users in cell b at time-slot t. Under the aforementioned assumptions, the baseband received signal model for user k in cell b, denoted as user k_b, can be expressed as

$$y_{k_b}(t) = \underbrace{\sqrt{\alpha_{k_b,b}}\,\mathbf{h}^{\mathsf{H}}_{k_b,b}(t)\mathbf{w}_{k_b}(t)d_{k_b}(t)}_{\text{desired signal}} + \underbrace{\sqrt{\alpha_{k_b,\bar{b}}}\,\mathbf{h}_{k_b,\bar{b}}(t)\sum_{j\in\mathcal{S}_{\bar{b}}(t)}\mathbf{w}_{j_{\bar{b}}}(t)d_{j_{\bar{b}}}(t)}_{\text{intercell interference}} + n_{k_b}(t), \quad (4.2)$$

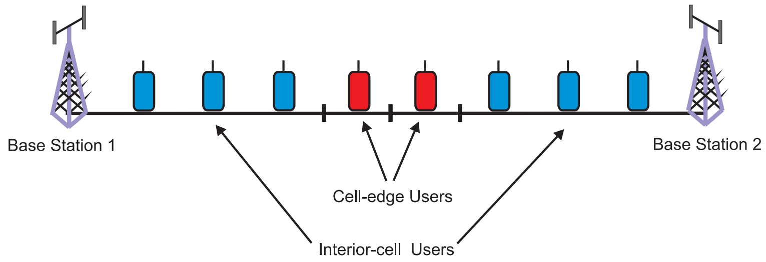


Figure 4.4: Schematic illustration of a two-cell MU-MIMO system.

where $\alpha_{k_b,b'}$ captures the pathloss between user k_b and base station b' and follows the model used in Section 3.2.2, $\mathbf{h}_{k_b,b'}(t)$ denotes the channel vector between user k_b and base station b', $\mathbf{w}_{k_b}(t)$ is the beamforming vector for user k_b at base station b, and $d_{k_b}(t)$ denotes the data symbol intended for user k_b. Here, we assume a per base station power constraint P, i.e., $\sum_{k\in\mathcal{S}_b(t)} \mathbb{E}[d^2_{k_b}] \leq P$, for $b = 1, 2$. Note that the intracell interference is essentially zero due to ZF beamforming inside each cell. The instantaneous achievable rate of user k_b assuming single-user detection is given by

$$R_k(t) = \log_2\left(1 + \frac{\alpha_{k_b,b}|\mathbf{h}^{\mathsf{H}}_{k_b,b}(t)\mathbf{w}_{k_b}(t)|^2 p_{k_b}}{1 + \alpha_{k_b,\bar{b}}\sum_{j\in\mathcal{S}_{\bar{b}}(t)} |\mathbf{h}^{\mathsf{H}}_{k_b,\bar{b}}(t)\mathbf{w}_{j_{\bar{b}}}(t)|^2 p_{j_{\bar{b}}}}\right). \tag{4.3}$$

We consider a simple coordination strategy that involves the cancelation of intercell interference to the cell-edge user in the neighboring cell if that user is scheduled for transmission by its home base station. To implement this strategy, at any given time-slot each base station performs a first step independent scheduling in its own cell and upon the selection of the cell-edge user, it informs the neighboring base station to update its beamformer in order to cancel the interference to this user. The neighboring base station then revise its scheduling, i.e., it tries to schedule a user in its own cell while canceling the interference to the scheduled cell-edge user in the other cell. One important issue that we would like to highlight here is that the base station at each cell cannot compute the instantaneous user rate at the scheduling stage as the intercell interference power is not known at this stage. This motivates the use of average intercell interference power to derive an estimate for the achievable rate [55]. The instantaneous rate estimate for user k_b can be written as

$$\tilde{R}_k(t) = \log_2\left(1 + \frac{\alpha_{k_b,b}|\mathbf{h}^{\mathsf{H}}_{k_b,b'}(t)\mathbf{w}_{k_b}(t)|^2 p_{k_b}}{1 + \alpha_{k_b,\bar{b}}P}\right). \tag{4.4}$$

This estimate rate will be used at each base station when performing PFS in (3.29). However, the actual achievable rate measured by the user is fed back to the base station to update the ergodic rates in (3.30).

Next, we consider three different systems for performance comparisons: 1) a baseline system, in which each base station independently selects and serves users in its own cell, denoted as the *conventional scheme*; 2) the coordinated system we just described with the instantaneous rate estimate calculated based on the average intercell interference power for all users, denoted as the *coordinated scheme (case I)*; 3) the same system as in 2, but with the intercell interference power of cell-edge user set to zero when computing the instantaneous rate estimate, denoted as the *coordinated scheme (case II)*. Note that this assumption relies on the described coordination strategy which forces the neighboring base station to cancel the interference to the cell-edge user in the other cell upon its selection.

Figure 4.5(a) shows the user throughput versus user locations for a cell-edge SNR of 10 dB. As can be seen, the coordinated scheme (case II) enhances the throughput for the cell-edge user at the expence of a relatively small loss for the cell-interior users. We also observe that although the intercell interference is completely canceled for the cell-edge user in coordinated scheme (case I), its throughput is increased slightly compared to the baseline scheme. To investigate the reason for this observation, we have plotted the net average rate and the activity fraction of the users in Fig. 4.5(b) and Fig. 4.5(c), respectively. Surprisingly, we observe that the net average rate of the cell-edge user in coordinated scheme (case I) is greater than that in coordinated scheme (case II). Note that the intercell interference for this user in both coordinated schemes is completely canceled. In addition, the channel statistics and beamforming strategy at its home cell are also the same in both schemes. As a result, this observation seems to be related to the power allocation. In fact, our simulation shows that in case I, the cell-edge user is more often selected as the only selected user in the cell (with full base station power allocated to it) than in case II. The activity fraction of this user in case I is, however, smaller than that in case II, resulting in the cell-edge user experiencing a lower throughput in this case. This can be attributed to the assumption on the intercell interference for this user in case I, which results in a low instantaneous rate estimate at the scheduling stage, thereby reducing the probability of selection of this user. Note that the ergodic rate values used to determine the scheduling weight of this user in PFS (see (3.29)) are the same in both schemes as these values are updated in (3.30) based on the achieved instantaneous rate. This example shows how the throughput of users in a cellular network depends on the coupled operation of both the MAC layer scheduling and the PHY layer beamforming.

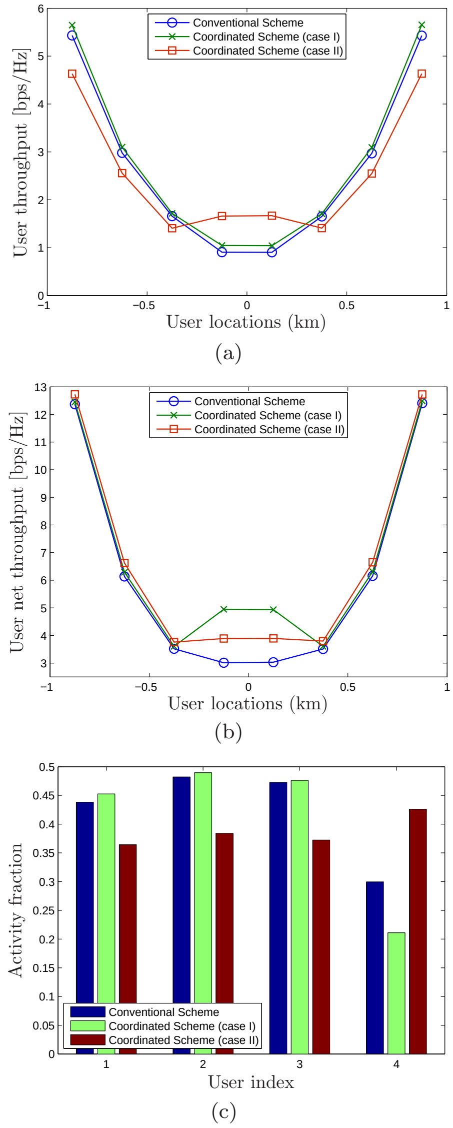


Figure 4.5: Comparision of throughput (a), net average rate (b), and activity fraction (c) for different systems.

4.4 Multicell Coordination Challenges

The performance gains provided by multicell coordination come at a cost which depends on the level of coordination. Here, we briefly explain some of the most important challenges that needs to be taken into account when designing such systems.

Synchronization

Coordinating base stations have to be synchronized in both time and frequency. The lack of synchronization in time results in inter-symbol interference, and in frequency leads to inter-carrier interference. Especially, different propagation delays of each user to the coordinating base stations might conflict with the guard interval, which limits the maximum distance between the coordinating base stations [56, 57].

Pilot and Feedback Overhead

Acquisition of user CSI to multiple base stations requires a larger number of orthogonal pilot sequences [4, 47, 58]. In FDD systems, base stations transmit orthogonal pilot sequences in the downlink to enable the users to estimate their channel from all coordinating base stations. In this case, the number of pilot sequences is proportional to the number of coordinating base stations' antennas, and is independent of the number of users in the system. Each user then feeds back its estimated channels to its serving base station over the feedback links. Hence, the load on the feedback links grows proportionally to the number of coordinating base stations. In TDD systems, however, user CSI is obtained at each base station using the orthogonal pilot sequences transmitted by the users. The number of pilot sequences is proportional to the number of users in the system, and is independent of the number of coordinating base stations and the number of antennas per base station. In order to demonstrate the amount of feedback overhead we express the following example.

EXAMPLE 2: Assume a user, equipped with 2 antennas, is going to communicate with 3 base stations, each with 4 antennas in a system that employs 256-tone OFDM. Furthermore, due to the mobility of the user there is a Doppler spread of 100 Hz. The channel estimation is assumed to happen 10 times faster than the Doppler. Using an 8 bits quantizer per real-valued coefficient the total feedback load for one user will be $8 \times 256 \times 48 \times 1000 = 93.3$ Mbps.

Backhaul Overhead

Data and CSI sharing over the backhaul requires high-capacity backhaul links. This makes joint processing a very complex scheme. Coordinated scheduling/beamforming is, however, of lower complexity as it only requires the sharing of CSI, which has been shown to occupy a very small fraction of the total backhaul bandwidth [59]. Furthermore, in practice the backhaul links are subject to delay [60] and error [61] and are of finite capacity [62,63], which needs to be taken into account when making a realistic design.

Complexity

Optimal centralized resource allocation including user scheduling, beamformer design, power and subcarrier allocation becomes prohibitive in coordinated systems even upon the availability of perfect CSI. This has motivated the development of distributed algorithms to reduce the complexity [64, 65].

Clustering

Network-wide coordination is extremely complicated to implement in practice. In fact, it has been shown that most of the network-wide coordination gain can be achieved by exploiting coordination among the neighboring base stations which cause the maximum interference to each other. The signal from farther base stations, would it be either desired or interference signal, is attenuated by the pathloss, while it imposes the same amount of overhead as the nearby base stations. This has motivated cluster-based coordination, in which coordinating clusters of limited size are selected in a static or dynamic fashion [66, 67].

Chapter 5

Purpose and Contributions

The purpose of the thesis is to investigate the design and the performance of spectrally-efficient coordinated MU-MIMO downlink systems. This chapter summarizes the contributions of the thesis, which are included in the form of five appended papers. Papers A and B consider multicell coordination in the form of network MIMO. The aim of these two papers is to develop analytical tools to enable the analysis of such systems instead of using intensive Monte-Carlo simulation. In particular, Paper A uses the developed analytical framework to propose a low-complexity scheduling algorithm with reduced feedback requirement. Papers C and D focus on multicell coordination in the form of CSB, which is of lower complexity compared to network MIMO, as there is no information data exchange among the base stations. In these two papers, the main goal is to propose low-complexity scheduling and beamforming strategies that require a limited amount of inter-base station CSI exchange. Finally, Paper E considers the coordination of the tilt angle selection at all base stations to enhance the throughput of multicell multiple-input single-output (MISO) systems. In the next section, we provide a short summary of these five appended papers. We then present some open research problems and future work in the field. Finally, a list of related publication by the author which are not included in the thesis is also presented.

5.1 Included Papers

Paper A - Multi-mode Transmission in Network MIMO Downlink with Incomplete CSI

This paper considers a network MIMO downlink system with multi-antenna base stations, which are connected to a central unit and communicate with multi-antenna users. In such a network, the overhead of obtaining perfect CSI of all users at the central unit to exploit opportunistic scheduling is daunting. The main purpose of the paper is to develop a low-complexity scheduling algorithm based only on the knowledge of the average received SNR at each user from all cooperating BSs, denoted as *incomplete CSI*. This is achieved by utilizing the results of random matrix theory, where an analytical framework is proposed to approximate the ergodic rate of each user with different number of data streams. Using these approximate ergodic rates, a joint user and mode selection algorithm is proposed, where only the scheduled users need to feed back instantaneous CSI. Simulation results demonstrate that the developed analytical framework provides a good approximation for a practical number of antennas.

Paper B - On the Achievable Ergodic Rate of Network MIMO Systems With Imperfect CSI

In this paper, we also consider the downlink of a network MIMO system, where multi-antenna base stations jointly serve multiple single-antenna users. Under realistic conditions of spatial user distribution and distance-dependent pathloss, the aggregate channel vector of each user to multiple base stations, denoted as the network MIMO channel vector, has *non-identically* distributed elements. This makes the mathematical treatment of such systems tedious as most of the available tools rely on channel vectors with independently and identically distributed (i.i.d.) elements. In this paper, using the statistical properties of Gamma random variables, we propose a new method to represent the network MIMO channel via an equivalent i.i.d. MIMO channel. We then use this method to derive an accurate analytical expression for the user ergodic rate assuming imperfect channel state information.

Paper C - Joint Scheduling and Intercell Interference Management in Multicell MISO Networks

Paper C investigates the analysis and performance of CSB as a lower-complexity coordination approach. We focus on the downlink of a clustered multicell network, in which each active single-antenna user receives information data from a single multi-antenna base station, while suffering co-channel interference from neighboring base stations. To manage the intercell interference within each cluster, coordination is employed via joint operation of user scheduling and downlink beamforming. We apply a stochastic optimization framework to find the optimal joint user scheduling and beamforming strategy. As the optimal solution has high complexity, a novel low-complexity algorithm is also proposed. Practical issues including inter-cluster interference suppression for cluster-edge users, and the effect of downlink training and imperfect channel knowledge are also investigated. Simulation results demonstrate that the proposed algorithm provides a significant throughput gain over the conventional non-coordinated system, especially at the cell-edge.

Paper D - Coordinated User Scheduling in the Multicell MIMO Downlink

Similar to Paper C, this paper also focuses on CSB in a two-cell multiuser MIMO network. A novel, distributed coordinated user scheduling (CUS) algorithm for intercell interference mitigation in the downlink is proposed. The intercell interference mitigation is achieved through the exchange of necessary CSI among the base stations, and the revision of the scheduling decisions and beamformer designs at each base station. Furthermore, mitigation of intercell interference is only performed for the cell-edge users, so that the amount of inter-base station signaling overhead is minimized. Our simulation results demonstrate that the proposed coordinated scheduling algorithm significantly improves the cell-edge users' throughput compared to conventional systems with only a negligible amount of CSI sharing among the base stations, and a relatively small throughput loss for the cell-interior users.

Paper E - Throughput Optimization in Multicell MISO Networks via Coordinated User-Specific Tilting

In Paper E, we focus on *antenna tilt*, which is an important antenna parameter whose impact in the context of coordinated systems has not been fully explored so far. We propose a novel framework to coordinatively select the tilt angles at all base stations. We further assume that the tilt angles can be remotely changed as fast as necessary by the network operator. Contrary to the conventional systems in which a fixed tilt angle is employed at each base station according to some statistical measures, in the proposed scheme the tilt angles are adjusted to the location of the scheduled user at each scheduling instance. Assuming the availability of location and channel statistics information of the scheduled users at all base stations, an accurate analytical expression for user ergodic rate is provided, which enables a decentralized deployment of the proposed framework at each BS. The superiority of the proposed coordinated tilting over the conventional schemes with a fixed tilt angle at each base station is shown via simulation.

5.2 Future Work

Throughout the thesis, we have made some simplifying assumptions which makes the obtained results serve as an upper bound for the actual performance. In practice, feedback links, assumed to be perfect here, are subject to both error and delay, which can severely degrade the performance. The backhaul links, used to share data and CSI among the base station, have also been assumed to be perfect in this thesis, which is not a valid assumption in practice. Similar to feedback links, backhaul links are also subject to delay and error. For example, the acquired CSI in the neighboring base stations used to perform scheduling, design precoders, etc., might not be correlated with the true CSI due to delay. Future work should also look into optimized methods for forming coordination clusters. Especially in coordination clusters with limited size, the optimal methods to deal with cluster-edge users, and the unknown inter-cluster interference have not been addressed thoroughly. Furthermore, as the future users are equipped with at least two antennas, the intercell interference mitigation can in part be performed at the user side using the interference alignment techniques [68].

5.3 Related Contributions

Other related publications by the author, which are not included in this thesis, are listed below.

- A. Wolfgang, **N. Seifi**, and T. Ottosson, " Resource Allocation and Linear Precoding for Relay Assisted Multiuser MIMO Systems", in *Proc. of International ITG Workshop on Smart Antennas*, Damstdart, Germany, Feb. 2008.
- **N. Seifi**, A. Wolfgang, and T. Ottosson, "Downlink Performance and Capacity of Distributed Antenna Systems Based on Realistic Channel Model", in *Proc. of International ITG Workshop on Smart Antennas*, Damstdart, Germany, Feb. 2008.
- **N. Seifi**, A. Wolfgang, and T. Ottosson, "Performance Analysis of Distributed versus Co-located MIMO-OFDM" *1st COST2100 Workshop on MIMO and Cooperative Communications*, Trondheim, Norway, May 2008.
- X. Wei, T. Weber, A. Wolfgang, and **N. Seifi**, "Joint Transmission with Significant CSI in the Downlink of Distributed Antenna Systems", in *Proc. of IEEE International Conference on Communications (ICC)*, Dresden, Germany, Jun. 2009.

- **N. Seifi**, T. Ottosson, M. Viberg, M. Coldrey, and A. Wolfgang, "An Efficient Signaling for Multimode Transmission in Multiuser MIMO" in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, USA, Mar. 2010.

- **N. Seifi**, M. Viberg, R. W. Heath Jr., J. Zhang, and M. Coldrey, "Coordinated Single-Cell vs Multi-Cell Transmission with Limited-Capacity Backhaul" in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, Nov. 2010.

- **N. Seifi**, M. Coldrey, M. Matthaiou, and M. Viberg "Impact of Base Station Antenna Tilt on the Performance of Network MIMO Systems" to appear in *Proc. IEEE Vehicular Technology Conference (VTC)*, Yokohama, Japan, May 2012.

- B. Makki, **N. Seifi**, and T. Eriksson, "Multiuser Diversity Using Two-Step Feedback", *accepted for publication in IET Communication, 2012.*

- **N. Seifi**, M. Coldrey, and M. Viberg "Decentralized Intercell Interference Management in Multicell MU-MIMO Networks" *in preparation.*

Bibliography

[1] S. Landström, A. Furuskär, K. Johansson, L. Falconetti, and F. Kronestedt, "Heterogeneous networks-increasing cellular capacity," *Ericsson Review 1*, 2011.

[2] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution, Second Edition: HSPA and LTE for Mobile Broadband*, Academic Press, 2008.

[3] G. J. Foschini, K. Karakayali, and R. A. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEE Proc. Commun.*, vol. 153, no. 4, pp. 548 – 555, Aug. 2006.

[4] G. Caire, S. A. Ramprashad, and H. C. Papadopoulos, "Rethinking network MIMO: Cost of CSIT, performance analysis, and architecture comparisons," in *Proc. of Info. Theory and Application Works. (ITA)*, May 2010.

[5] J. Zhang and J. Andrews, "Adaptive spatial intercell interference cancellation in multicell wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1455–1468, Dec. 2010.

[6] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 2001.

[7] A. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, NY, USA, 2005.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, New York, NY, USA, 2005.

[9] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proc. of IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[10] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, 1948.

[11] F. Khan, *LTE for 4G mobile broadband: air interface technologies and performance*, Cambridge University Press, 2009.

[12] S. W. Halpern, "Reuse partitioning in cellular systems," in *Proc. of 33rd IEEE Veh. Technol. Conf. (VTC)*, May 1983, vol. 33.

[13] Y. Liang, A. Goldsmith, G. Foschini, R. Valenzuela, and D. Chizhik, "Evolution of base stations in cellular networks: Denser deployment versus coordination," in *Proc. of IEEE Int. Conf. Commun. (ICC)*, May 2008.

[14] J. Winters, "On the capacity of radio communication systems with diversity in a rayleigh fading environment," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 5, pp. 871–878, Jun. 1987.

[15] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41 – 59, 1996.

[16] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.

[17] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[18] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[19] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Info. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[20] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Info. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[21] John Proakis, *Digital Communications*, McGraw-Hill Science/Engineering/Math, 4 edition, Aug. 2000.

[22] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Trans. Info. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

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

[24] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when usingmultiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

[25] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[26] C. B. Peel, Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Commun. Mag.*, pp. 60–67, Oct. 2004.

[27] D. Gesbert, M. Kountouris, R. W. Heath Jr., C. B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[28] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Info. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[29] M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[30] W. Yu and W. Rhee, "Degrees of freedom in wireless multiuser spatial multiplex systems with multiple antennas," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1747–1753, Oct. 2006.

[31] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[32] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[33] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Info. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[34] S. Shim, J. Kwak, R. W. Heath Jr., and J. Andrews, "Block diagonalization for multi-user MIMO with other-cell interference," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2671–2681, Jul. 2008.

[35] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[36] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[37] M. J. Neely, E. Modiano, and C. E. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 138–152, Feb. 2003.

[38] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Info. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[39] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3857–3868, Oct. 2005.

[40] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[41] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.

[42] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Info. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[43] B. Makki and T. Eriksson, "On hybrid ARQ and quantized CSI feedback schemes in quasi-static fading channels," *in press IEEE Trans. Commun.*, 2012.

[44] B. Makki and T. Eriksson, "On the average rate of quasi-static fading channels with ARQ and CSI feedback," *IEEE Commun. Letters*, vol. 14, no. 9, pp. 806–808, Sept. 2010.

[45] B. Makki and T. Eriksson, "CSI feedback in correlated slow-fading channels," *IEEE Commun. Letters*, vol. 15, no. 12, pp. 1294–1297, Dec. 2011.

[46] B. Makki and T. Eriksson, "On the average rate of HARQ-based quasi-static spectrum sharing networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 65–77, Jan. 2012.

[47] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.

[48] S. Wagner, R. Couillet, D. T. M. Slock, and M. Debbah, "Large system analysis of zero-forcing precoding in MISO broadcast channels with limited feedback," *Proc. IEEE Int. Works. Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2010.

[49] U. Salim, D. Gesbert, and D. Slock, "Combining training and quantized feedback in multi-antenna reciprocal channels," *Submitted to IEEE Trans. Signal Process.*, Apr. 2010.

[50] J. G. Andrews, W. Choi, and R. W. Heath Jr., "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Commun. Mag.*, Dec. 2007.

[51] H. Holma and A. Toskala, Eds., *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 2001.

[52] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[53] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced," *IEEE Wireless Commun. Mag.*, vol. 17, no. 3, pp. 26–34, Jun. 2010.

[54] Rui Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.

[55] S. A. Ramprashad, G. Caire, and H. C. Papadopoulos, "A joint scheduling and cell clustering scheme for MU-MIMO downlink with limited coordination," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2010.

[56] B. W. Zarikoff and J .K. Cavers, "Coordinated multi-cell systems: Carrier frequency offset estimation and correction," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1490–1501, Dec. 2010.

[57] H. Zhang, N. B. Mehta, A. F. Molisch, J. Zhang, and H. Dai, "Asynchronous interference mitigation in cooperative base station systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 155–165, Jan. 2008.

[58] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation and reduced-complexity scheduling," *in press IEEE Trans. Info. Theory*, 2010.

[59] D. Samardzija and H. Huang, "Determining backhaul bandwidth requirements for network MIMO," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Aug. 2009.

[60] R. Bhagavatula and R. W. Heath Jr., "Adaptive bit partitioning for multicell intercell interference nulling with delayed limited feedback," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3824–3836, Aug. 2011.

[61] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, Jun. 2011.

[62] R. Zakhour and D. Gesbert, "Optimized data sharing in multicell MIMO with finite backhaul capacity," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6102–6111, Dec. 2011.

[63] P. Marsch and G. Fettweis, "Uplink comp under a constrained backhaul and imperfect channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1730–1742, Jun. 2011.

[64] A. Tölli, H. Pennanen, and P. Komulainen, "Distributed coordinated multi-cell transmission based on dual decomposition," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2009.

[65] R. Zakhour, Z. Ho, and D. Gesbert, "Distributed beamforming coordination in multicell MIMO channels," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Apr. 2009.

[66] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[67] J. Gong, S. Zhou, Z. Niu, L. Geng, and M. Zheng, "Joint scheduling and dynamic clustering in downlink cellular networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011.

[68] C. Suh, M. Ho, and D. N. C. Tse, "Downlink interference alignment," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2616–2626, Sep. 2011.

[69] R. Irmer and et al., "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.