Linköping Studies in Science and Technology
Dissertations, No. 1642

Massive MIMO: Fundamentals and System Designs

Hien Quoc Ngo

Division of Communication Systems
Department of Electrical Engineering (ISY)
Linköping University, SE-581 83 Linköping, Sweden
www.commsys.isy.liu.se

Linköping 2015

Massive MIMO: Fundamentals and System Designs

ISBN 978-91-7519-147-8
ISSN 0345-7524
Printed in Sweden by LiU-Tryck, Linköping 2015

Cảm ơn gia đình tôi, cảm ơn Em,

vì đã luôn bên cạnh tôi.

Abstract

The last ten years have seen a massive growth in the number of connected wireless devices. Billions of devices are connected and managed by wireless networks. At the same time, each device needs a high throughput to support applications such as voice, real-time video, movies, and games. Demands for wireless throughput and the number of wireless devices will always increase. In addition, there is a growing concern about energy consumption of wireless communication systems. Thus, future wireless systems have to satisfy three main requirements: i) having a high throughput; ii) simultaneously serving many users; and iii) having less energy consumption. Massive multiple-input multiple-output (MIMO) technology, where a base station (BS) equipped with very large number of antennas (collocated or distributed) serves many users in the same time-frequency resource, can meet the above requirements, and hence, it is a promising candidate technology for next generations of wireless systems. With massive antenna arrays at the BS, for most propagation environments, the channels become favorable, i.e., the channel vectors between the users and the BS are (nearly) pairwisely orthogonal, and hence, linear processing is nearly optimal. A huge throughput and energy efficiency can be achieved due to the multiplexing gain and the array gain. In particular, with a simple power control scheme, Massive MIMO can offer uniformly good service for all users. In this dissertation, we focus on the performance of Massive MIMO. The dissertation consists of two main parts: fundamentals and system designs of Massive MIMO.

In the first part, we focus on fundamental limits of the system performance under practical constraints such as low complexity processing, limited length of each coherence interval, intercell interference, and finite-dimensional channels. We first study the potential for power savings of the Massive MIMO uplink with maximum-ratio combining (MRC), zero-forcing, and minimum mean-square error receivers, under perfect and imperfect channels. The energy and spectral efficiency tradeoff is investigated. Secondly, we consider a physical channel model where the angular domain is divided into a finite number of distinct directions. A lower bound on the capacity is derived, and the effect of pilot contamination in this finite-dimensional channel model is analyzed. Finally, some aspects of favorable propagation in Massive MIMO under Rayleigh fading and line-of-sight (LoS) channels are investigated. We show that both Rayleigh fading and LoS environments offer favorable propagation.

In the second part, based on the fundamental analysis in the first part, we propose some system designs for Massive MIMO. The acquisition of channel state information (CSI) is very important in Massive MIMO. Typically, the channels are estimated at the BS through uplink training. Owing to the limited length of the coherence interval, the system performance is limited by pilot contamination. To reduce the pilot contamination effect, we propose an eigenvalue-decomposition-based scheme to estimate the channel directly from the received data. The proposed scheme results in better performance compared with the conventional training schemes due to the reduced pilot contamination. Another important issue of CSI acquisition in Massive MIMO is how to acquire CSI at the users. To address this issue, we propose two channel estimation schemes at the users: i) a downlink "beamforming training" scheme, and ii) a method for blind estimation of the effective downlink channel gains. In both schemes, the channel estimation overhead is independent of the number of BS antennas. We also derive the optimal pilot and data powers as well as the training duration allocation to maximize the sum spectral efficiency of the Massive MIMO uplink with MRC receivers, for a given total energy budget spent in a coherence interval. Finally, applications of Massive MIMO in relay channels are proposed and analyzed. Specifically, we consider multi-pair relaying systems where many sources simultaneously communicate with many destinations in the same time-frequency resource with the help of a Massive MIMO relay. A Massive MIMO relay is equipped with many collocated or distributed antennas. We consider different duplexing modes (full-duplex and half-duplex) and different relaying protocols (amplify-and-forward, decode-and-forward, two-way relaying, and one-way relaying) at the relay. The potential benefits of massive MIMO technology in these relaying systems are explored in terms of spectral efficiency and power efficiency.

Populärvetenskaplig Sammanfattning

Det har skett en massiv tillväxt av antalet trådlöst kommunicerande enheter de senaste tio åren. Idag är miljarder av enheter anslutna och styrda över trådlösa nätverk. Samtidigt kräver varje enhet en hög datatakt för att stödja sina applikationer, som röstkommunikation, realtidsvideo, film och spel. Efterfrågan på trådlös datatakt och antalet trådlösa enheter kommer alltid att tillta. Samtidigt kan inte strömförbrukningen hos de trådlösa kommunikationssystemen tillåtas att öka. Således måste framtida trådlösa kommunikationssystem uppfylla tre huvudkrav: i) hög datatakt ii) kunna betjäna många användare samtidigt iii) lägre strömförbrukning.

Massiv MIMO ("multiple-input multiple output"), en teknik där basstationen är utrustad med ett stort antal antenner och samtidigt betjänar många användare över samma tid-frekvensresurs, kan uppfylla ovanstående krav. Följaktligen kan det betraktas som en lovande kandidat för nästa generations trådlösa system. För de flesta utbredningsmiljöer blir kanalen fördelaktig med en massiv antennuppställning (en uppställning av, låt säga, hundra antenner eller fler), det vill säga kanalvektorerna mellan användare och basstation blir (nästan) parvis ortogonala, vilket gör linjär signalbehandling nästan optimal. Den höga datatakten och låga strömförbrukningen kan åstadkommas tack vare multiplexeringsvinsten och antennförstärkningen. I synnerhet kan massiv MIMO erbjuda en likformigt bra betjäning av alla användare med en enkel effektallokeringsmetod.

I denna avhandling börjar vi med att fokusera på grunderna av massiv MIMO. Speciellt kommer vi att studera de grundläggande begränsningarna av systemets prestanda i termer av spektral effektivitet och energieffektivitet när massiva antennuppställningar används. Detta kommer vi att göra med beaktande av praktiska begränsningar hos systemet, som lågkomplexitetsbehandling (till exempel linjär behandling av signaler), begränsad längd av varje koherensinterval, ofullständig kanalkännedom, intercell-interferens och ändlig-dimensionella kanaler. Dessutom undersöks några aspekter hos fördelaktig utbredning i massiv MIMO med rayleighfädning och kanaler med rakt sikt. Baserat på dessa grundläggande analyser föreslår vi sedan några systemkonstruktioner för massiv MIMO. Mer precist föreslår vi några

metoder för kanalskattning både för basstationen och för användarna, vilka ämnar minimera effekten av pilotkontaminering och kanalovisshet. Den optimala pilot- och dataeffekten så väl som valet av längden av träningsperioden studeras. Till slut föreslås och analyseras användandet av massiv MIMO i reläkanaler.

Acknowledgments

I would like to extend my sincere thanks to my supervisor, Prof. Erik G. Larsson, for his valuable support and supervision. His advice, guidance, encouragement, and inspiration have been invaluable over the years. Prof. Larsson always keeps an open mind in every academic discussion. I admire his critical eye for important research topics. I still remember when I began my doctoral studies, Prof. Larsson showed me the first paper on Massive MIMO and stimulated my interest for this topic. This thesis would not have been completed without his guidance and support.

I would like to thank Dr. Thomas L. Marzetta at Bell Laboratories, Alcatel-Lucent, USA, for his cooperative work, and for giving me a great opportunity to join his research group as a visiting scholar. It has been a great privilege to be a part of his research team. He gave me valuable help whenever I asked for assistance. I have learnt many useful things from him. I would also like to thank Dr. Alexei Ashikhmin and Dr. Hong Yang for making my visit at Bell Laboratories, Alcatel-Lucent in Murray Hill such a great experience.

I was lucky to meet many experts in the field. I am thankful to Dr. Michail Matthaiou at Queen's University Belfast, U.K., for his great cooperation. I have learnt a lot from his maturity and expertise. Many thanks to Dr. Trung Q. Duong at Queen's University Belfast, U.K., and Dr. Himal A. Suraweera at University of Peradeniya, Sri Lanka, for both technical and non-technical issues during the cooperative work. I would like to thank Dr. Le-Nam Tran at Maynooth University, Ireland, for his explanations and discussions on the optimization problems which helped me a lot. I am also thankful to all of my co-authors for the collaboration over these years: Dr. G. C. Alexandropoulos (France Research Center, Huawei Technologies Co. Ltd.), Prof. H-J. Zepernick (Blekinge Institute of Technology, Sweden), Dr. C. Yuen (Singapore University of Technology and Design, Singapore), Dr. A. K. Papazafeiropoulos (Imperial College, U.K.), Dr. H. Phan (University of Reading, U.K.), Dr. M. Elkashlan (Queen Mary University of London, U.K.), and Mr. L. Wang (Queen Mary University of London, U.K.).

The warmest thank to my colleagues at Communication Systems, ISY, Linköping University, for the stimulating discussions, and for providing the fun environment in which we have learnt and grown during the past 4+ years. Special thanks to my fellow PhD students: Chaitanya, Reza, Mirsad, Johannes, Antonios, Erik Axell, Victor, Christopher, and Marcus.

Finally, I would like to thank my family and friends, for their constant love, encouragement, and limitless support throughout my life.

Linköping, January 2015
Hien Quoc Ngo

Abbreviations

AF	Amplify-and-Forward
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
CDF	Cumulative Distribution Function
CSI	Channel State Information
DF	Decode-and-Forward
DL	Downlink
DPC	Dirty Paper Coding
EVD	Eigenvalue Decomposition
FD	Full Duplex
FDD	Frequency Division Duplexing
HD	Half Duplex
i.i.d.	Independent and Identically Distributed
ILSP	Iterative Least-Square with Projection
LDPC	Low-Density Parity-Check
LTE	Long Term Evolution
LoS	Line-of-Sight
LS	Least-Squares
MAC	Multiple-Access Channel
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MMSE	Minimum Mean Square Error
MSE	Mean-Square Error
ML	Maximum Likelihood
MRC	Maximum Ratio Combining
MRT	Maximum Ratio Transmission
MU-MIMO	Multiuser MIMO

PDF	Probability Density Function
OFDM	Orthogonal Frequency Division Multiplexing
QAM	Quadrature Amplitude Modulation
RV	Random Variable
SEP	Symbol Error Probability
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SIR	Signal-to-Interference Ratio
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
TDD	Time Division Duplexing
TWRC	Two-Way Relay Channel
UL	Uplink
ZF	Zero-Forcing

Contents

Part I

Introduction

Chapter 1

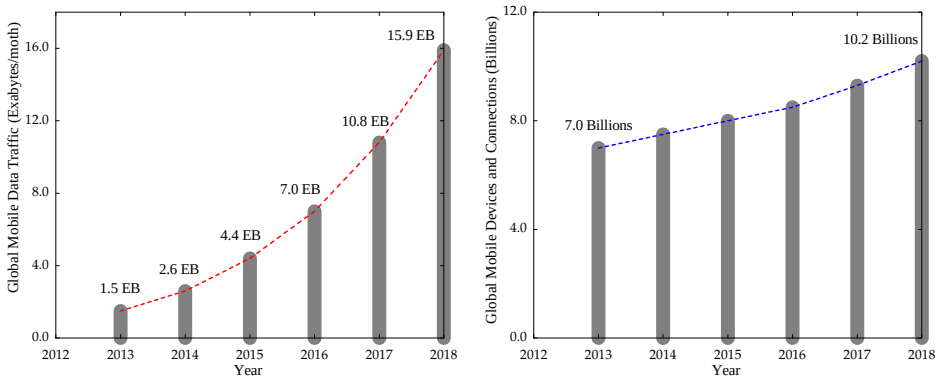
Motivation

During the last years, data traffic (both mobile and fixed) has grown exponentially due to the dramatic growth of smartphones, tablets, laptops, and many other wireless data consuming devices. The demand for wireless data traffic will be even more in future [1–3]. Figures 1.1 shows the demand for mobile data traffic and the number of connected devices. Global mobile data traffic is expected to increase to 15.9 exabytes per month by 2018, which is about an 6-fold increase over 2014. In addition, the number of mobile devices and connections are expected to grow to 10.2 billion by 2018. New technologies are required to meet this demand. Related to wireless data traffic, the key parameter to consider is wireless throughput (bits/s) which is defined as:

$$\text{Throughput} = \text{Bandwidth (Hz)} \times \text{Spectral efficiency (bits/s/Hz)}.$$

Clearly, to improve the throughput, some new technologies which can increase the bandwidth or the spectral efficiency or both should be exploited. In this thesis, we focus on techniques which improve the spectral efficiency. A well-known way to increase the spectral efficiency is using multiple antennas at the transceivers.

In wireless communication, the transmitted signals are being attenuated by fading due to multipath propagation and by shadowing due to large obstacles between the transmitter and the receiver, yielding a fundamental challenge for reliable communication. Transmission with multiple-input multiple-output (MIMO) antennas is a well-known diversity technique to enhance the reliability of the communication. Furthermore, with multiple antennas, multiple streams can be sent out and hence, we can obtain a multiplexing gain which significantly improves the communication capacity. MIMO systems have gained significant attention for the past decades, and are now being incorporated into several new generation wireless standards (e.g., LTE-Advanced, 802.16m).



(a) Global mobile data traffic.

(b) Global mobile devices and connections growth.

Figure 1.1: Demand for mobile data traffic and number of connected devices. (Source: Cisco [3])

The effort to exploit the spatial multiplexing gain has been shifted from MIMO to multiuser MIMO (MU-MIMO), where several users are simultaneously served by a multiple-antenna base station (BS). With MU-MIMO setups, a spatial multiplexing gain can be achieved even if each user has a single antenna [4]. This is important since users cannot support many antennas due to the small physical size and low-cost requirements of the terminals, whereas the BS can support many antennas. MU-MIMO does not only reap all benefits of MIMO systems, but also overcomes most of propagation limitations in MIMO such as ill-behaved channels. Specifically, by using scheduling schemes, we can reduce the limitations of ill-behaved channels. Line-of-sight propagation, which causes significant reduction of the performance of MIMO systems, is no longer a problem in MU-MIMO systems. Thus, MU-MIMO has attracted substantial interest [4–9].

There always exists a tradeoff between the system performance and the implementation complexity. The advantages of MU-MIMO come at a price:

- Multiuser interference: the performance of a given user may significantly degrade due to the interference from other users. To tackle this problem, interference reduction or cancellation techniques, such as maximum likelihood multiuser detection for the uplink [10], dirty paper coding (DPC) techniques for the downlink [11], or interference alignment [12], should be used. These techniques are complicated and have high computational complexity.
- Acquisition of channel state information: in order to achieve a high spatial multiplexing gain, the BS needs to process the received signals coherently. This requires accurate and timely acquisition of channel state information (CSI). This can be challenging, especially in high mobility scenarios.

- User scheduling: since several users are served on the same time-frequency resource, scheduling schemes which optimally select the group of users depending on the precoding/detection schemes, CSI knowledge etc., should be considered. This increases the cost of the system implementation.

The more antennas the BS is equipped with, the more degrees of freedom are offered and hence, more users can simultaneously communicate in the same time-frequency resource. As a result, a huge sum throughput can be obtained. With large antenna arrays, conventional signal processing techniques (e.g. maximum likelihood detection) become prohibitively complex due to the high signal dimensions. The main question is whether we can obtain the huge multiplexing gain with low-complexity signal processing and low-cost hardware implementation.

In [13], Marzetta showed that the use of an excessive number of BS antennas compared with the number of active users makes simple linear processing nearly optimal. More precisely, even with simple maximum-ratio combining (MRC) in the uplink or maximum-ratio transmission (MRT) in the downlink, the effects of fast fading, intracell interference, and uncorrelated noise tend to disappear as the number of BS station antennas grows large. MU-MIMO systems, where a BS with a hundred or more antennas simultaneously serves tens (or more) of users in the same time-frequency resource, are known as *Massive MIMO* systems (also called very large MU-MIMO, hyper-MIMO, or full-dimension MIMO systems). In Massive MIMO, it is expected that each antenna would be contained in an inexpensive module with simple processing and a low-power amplifier. The main benefits of Massive MIMO systems are:

(1) *Huge spectral efficiency and high communication reliability*: Massive MIMO inherits all gains from conventional MU-MIMO, i.e., with M-antenna BS and K single-antenna users, we can achieve a diversity of order M and a multiplexing gain of $\min(M, K)$. By increasing both M and K, we can obtain a huge spectral efficiency and very high communication reliability.

(2) *High energy efficiency*: In the uplink Massive MIMO, coherent combining can achieve a very high array gain which allows for substantial reduction in the transmit power of each user. In the downlink, the BS can focus the energy into the spatial directions where the terminals are located. As a result, with massive antenna arrays, the radiated power can be reduced by an order of magnitude, or more, and hence, we can obtain high energy efficiency. For a fixed number of users, by doubling the number of BS antennas, while reducing the transmit power by two, we can maintain the original the spectral efficiency, and hence, the radiated energy efficiency is doubled.

(3) *Simple signal processing*: For most propagation environments, the use of an excessive number of BS antennas over the number of users yields favorable propagation where the channel vectors between the users and the BS are

pairwisely (nearly) orthogonal. Under favorable propagation, the effect of interuser interference and noise can be eliminated with simple linear signal processing (liner precoding in the downlink and linear decoding in the uplink). As a result, simple linear processing schemes are nearly optimal. Another key property of Massive MIMO is channel hardening. Under some conditions, when the number of BS antennas is large, the channel becomes (nearly) deterministic, and hence, the effect of small-scale fading is averaged out. The system scheduling, power control, etc., can be done over the large-scale fading time scale instead of over the small-scale fading time scale. This simplifies the signal processing significantly.

Massive MIMO is a promising candidate technology for next-generation wireless systems. Recently, there has been a great deal of interest in this technology [14–18]. Although there is much research work on this topic, a number of issues still need to be tackled before reducing Massive MIMO to practice [19–26].

Inspired by the above discussion, in this dissertation, we study the fundamentals of Massive MIMO including favorable propagation aspects, spectral and energy efficiency, and effects of finite-dimensional channel models. Capacity bounds are derived and analysed under practical constraints such as low-complexity processing, imperfect CSI, and intercell interference. Based on the fundamental analysis of Massive MIMO, resource allocation as well as system designs are also proposed.

In the following, brief introductions to multiuser MIMO and Massive MIMO are given in Chapter 2 and Chapter 3, respectively. In Chapter 4, we provide some mathematical preliminaries which will be used throughout the thesis. In Chapter 5, we list the specific contributions of the thesis together with a short description of the included papers. Finally, future research directions are discussed in Chapter 6.

Chapter 2

Mutiuser MIMO Cellular Systems

Massive MIMO is a MU-MIMO cellular system where the number of BS antennas and the number users are large. In this section, we will provide the basic background of MU-MIMO cellular systems in terms of communication schemes and signal detection, for both the uplink and downlink. For the sake of simplicity, we limit our discussions to the single-cell systems.

2.1 System Models and Assumptions

We consider a MU-MIMO system which consists of one BS and K active users. The BS is equipped with M antennas, while each user has a single-antenna. In general, each user can be equipped with multiple antennas. However, for simplicity of the analysis, we limit ourselves to systems with single-antenna users. See Figure 2.1. We assume that all K users share the same time-frequency resource. Furthermore, we assume that the BS and the users have perfect CSI. The channels are acquired at the BS and the users during the training phase. The specific training schemes depend on the system protocols (frequency-division duplex (FDD) or time-division duplex (TDD)), and will be discussed in detail in Section 2.5.

Let $\boldsymbol{H} \in \mathbb{C}^{M \times K}$ be the channel matrix between the K users and the BS antenna array, where the kth column of $\boldsymbol{H}$, denoted by $\boldsymbol{h}_k$, represents the $M \times 1$ channel vector between the kth user and the BS. In general, the propagation channel is modeled via large-scale fading and small-scale fading. But in this chapter, we ignore large-scale fading, and further assume that the elements of $\boldsymbol{H}$ are i.i.d. Gaussian distributed with zero mean and unit variance.

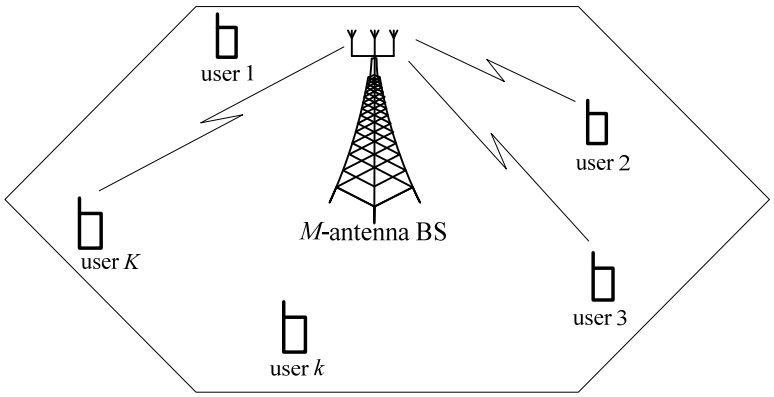


Figure 2.1: Multiuser MIMO Systems. Here, K single-antenna users are served by the M-antenna BS in the same time-frequency resource.

2.2 Uplink Transmission

Uplink (or reverse link) transmission is the scenario where the K users transmit signals to the BS. Let s_k, where $\mathbb{E}\left\{|s_k|^2\right\} = 1$, be the signal transmitted from the kth user. Since K users share the same time-frequency resource, the $M \times 1$ received signal vector at the BS is the combination of all signals transmitted from all K users:

$$\boldsymbol{y}_{\mathrm{ul}} = \sqrt{p_{\mathrm{u}}} \sum_{k=1}^{K} \boldsymbol{h}_k s_k + \boldsymbol{n} \tag{2.1}$$

$$= \sqrt{p_{\mathrm{u}}} \boldsymbol{H} \boldsymbol{s} + \boldsymbol{n}, \tag{2.2}$$

where p_{u} is the average signal-to-noise ratio (SNR), $\boldsymbol{n} \in \mathbb{C}^{M \times 1}$ is the additive noise vector, and $\boldsymbol{s} \triangleq [s_1 \ ... \ s_K]^T$. We assume that the elements of $\boldsymbol{n}$ are i.i.d. Gaussian random variables (RVs) with zero mean and unit variance, and independent of $\boldsymbol{H}$.

From the received signal vector $\boldsymbol{y}_{\mathrm{ul}}$ together with knowledge of the CSI, the BS will coherently detect the signals transmitted from the K users. The channel model (2.2) is the multiple-access channel which has the sum-capacity [27]

$$C_{\mathrm{ul,sum}} = \log_2 \det \left(\boldsymbol{I}_K + p_{\mathrm{u}} \boldsymbol{H}^H \boldsymbol{H} \right). \tag{2.3}$$

The aforementioned sum-capacity can be achieved by using the successive interference cancellation (SIC) technique [28]. With SIC, after one user is detected, its signal is subtracted from the received signal before the next user is detected.

2.3 Downlink Transmission

Downlink (or forward link) is the scenario where the BS transmits signals to all K users. Let $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$, where $\mathbb{E}\left\{\|\boldsymbol{x}\|^2\right\} = 1$, be the signal vector transmitted from the BS antenna array. Then, the received signal at the kth user is given by

$$y_{\mathrm{dl},k} = \sqrt{p_{\mathrm{d}}}\boldsymbol{h}_k^T\boldsymbol{x} + z_k, \tag{2.4}$$

where p_{d} is the average SNR and z_k is the additive noise at the kth user. We assume that z_k is Gaussian distributed with zero mean and unit variance. Collectively, the received signal vector of the K users can be written as

$$\boldsymbol{y}_{\mathrm{dl}} = \sqrt{p_{\mathrm{d}}}\boldsymbol{H}^T\boldsymbol{x} + \boldsymbol{z}, \tag{2.5}$$

where $\boldsymbol{y}_{\mathrm{dl}} \triangleq [y_{\mathrm{dl},1}\ y_{\mathrm{dl},2}\ \ldots\ y_{\mathrm{dl},K}]^T$ and $\boldsymbol{z} \triangleq [z_1\ z_2\ \ldots\ z_K]^T$. The channel model (2.5) is the broadcast channel whose sum-capacity is known to be

$$C_{\mathrm{sum}} = \max_{\substack{\{q_k\} \\ q_k\geq 0,\ \sum_{k=1}^{K} q_k \leq 1}} \log_2 \det\left(\boldsymbol{I}_M + p_{\mathrm{d}}\boldsymbol{H}^*\boldsymbol{D}_{\boldsymbol{q}}\boldsymbol{H}^T\right), \tag{2.6}$$

where $\boldsymbol{D}_{\boldsymbol{q}}$ is the diagonal matrix whose kth diagonal element is q_k. The sum-capacity (2.6) can be achieved by using the dirty-paper coding (DPC) technique.

2.4 Linear Processing

To obtain optimal performance, complex signal processing techniques must be implemented. For example, in the uplink, the maximum-likelihood (ML) multiuser detection can be used. With ML multiuser detection, the BS has to search all possible transmitted signal vectors $\boldsymbol{s}$, and choose the best one as follows:

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}\in\mathcal{S}^K} \|\boldsymbol{y}_{\mathrm{ul}} - \sqrt{p_{\mathrm{u}}}\boldsymbol{H}\boldsymbol{s}\|^2 \tag{2.7}$$

where $\mathcal{S}$ is the finite alphabet of s_k, $k = 1, 2, ..., K$. The problem (2.7) is a least-squares (LS) problem with a finite-alphabet constraint. The BS has to search over $|\mathcal{S}|^K$ vectors, where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$. Therefore, ML has a complexity which is exponential in the number of users.

The BS can use linear processing schemes (linear receivers in the uplink and linear precoders in the downlink) to reduce the signal processing complexity. These schemes are not optimal. However, when the number of BS antennas is large, it is shown in [13, 14] that linear processing is nearly-optimal. Therefore, in this thesis, we will consider linear processing. The details of linear processing techniques are presented in the following sections.

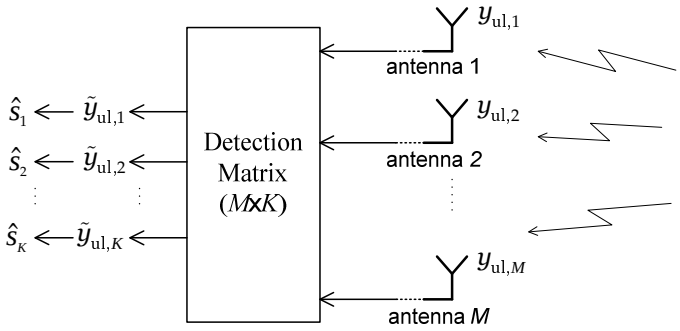


Figure 2.2: Block diagram of linear detection at the BS.

2.4.1 Linear Receivers (in the Uplink)

With linear detection schemes at the BS, the received signal $\boldsymbol{y}_{\text{ul}}$ is separated into K streams by multiplying it with an $M \times K$ linear detection matrix, $\boldsymbol{A}$:

$$\tilde{\boldsymbol{y}}_{\text{ul}} = \boldsymbol{A}^H \boldsymbol{y}_{\text{ul}} = \sqrt{p_{\text{u}}} \boldsymbol{A}^H \boldsymbol{H} \boldsymbol{s} + \boldsymbol{A}^H \boldsymbol{n}. \tag{2.8}$$

Each stream is then decoded independently. See Figure 2.2. The complexity is on the order of $K|\mathcal{S}|$. From (2.8), the kth stream (element) of $\tilde{\boldsymbol{y}}_{\text{ul}}$, which is used to decode s_k, is given by

$$\tilde{y}_{\text{ul},k} = \underbrace{\sqrt{p_{\text{u}}} \boldsymbol{a}_k^H \boldsymbol{h}_k s_k}_{\text{desired signal}} + \underbrace{\sqrt{p_{\text{u}}} \sum_{k' \neq k}^{K} \boldsymbol{a}_k^H \boldsymbol{h}_{k'} s_{k'}}_{\text{interuser interference}} + \underbrace{\boldsymbol{a}_k^H \boldsymbol{n}}_{\text{noise}}, \tag{2.9}$$

where $\boldsymbol{a}_k$ denotes the kth column of $\boldsymbol{A}$. The interference plus noise is treated as effective noise, and hence, the received signal-to-interference-plus-noise ratio (SINR) of the kth stream is given by

$$\texttt{SINR}_k = \frac{p_{\text{u}} |\boldsymbol{a}^H \boldsymbol{h}_k|^2}{p_{\text{u}} \sum_{k' \neq k}^{K} |\boldsymbol{a}_k^H \boldsymbol{h}_{k'}|^2 + \|\boldsymbol{a}_k\|^2}. \tag{2.10}$$

We now review some conventional linear multiuser receivers.

a) *Maximum-Ratio Combining receiver:*

With MRC, the BS aims to maximize the received signal-to-noise ratio (SNR) of each stream, ignoring the effect of multiuser interference. From (2.9), the

kth column of the MRC receiver matrix $\boldsymbol{A}$ is:

$$\begin{aligned}\boldsymbol{a}_{\mathtt{mrc},k} &= \underset{\boldsymbol{a}_k \in \mathbb{C}^{M\times 1}}{\operatorname{argmax}} \frac{\text{power(desired signal)}}{\text{power(noise)}} \\ &= \underset{\boldsymbol{a}_k \in \mathbb{C}^{M\times 1}}{\operatorname{argmax}} \frac{p_{\mathrm{u}}|\boldsymbol{a}_k^H \boldsymbol{h}_k|^2}{\|\boldsymbol{a}_k\|^2}. \end{aligned} \tag{2.11}$$

Since

$$\frac{p_{\mathrm{u}}|\boldsymbol{a}_k^H \boldsymbol{h}_k|^2}{\|\boldsymbol{a}_k\|^2} \leq \frac{p_{\mathrm{u}}\|\boldsymbol{a}_k\|^2\|\boldsymbol{h}_k\|^2}{\|\boldsymbol{a}_k\|^2} = p_{\mathrm{u}}\|\boldsymbol{h}_k\|^2,$$

and equality holds when $\boldsymbol{a}_k = \text{const} \cdot \boldsymbol{h}_k$, the MRC receiver is: $\boldsymbol{a}_{\mathtt{mrc},k} = \text{const} \cdot \boldsymbol{h}_k$. Plugging $\boldsymbol{a}_{\mathtt{mrc},k}$ into (2.10), the received SINR of the kth stream for MRC is given by

$$\mathtt{SINR}_{\mathtt{mrc},k} = \frac{p_{\mathrm{u}}\|\boldsymbol{h}_k\|^4}{p_{\mathrm{u}}\sum_{k'\neq k}^{K}|\boldsymbol{h}_k^H \boldsymbol{h}_{k'}|^2 + \|\boldsymbol{h}_k\|^2} \tag{2.12}$$

$$\rightarrow \frac{\|\boldsymbol{h}_k\|^4}{\sum_{k'\neq k}^{K}|\boldsymbol{h}_k^H \boldsymbol{h}_{k'}|^2}, \text{ as } p_{\mathrm{u}} \rightarrow \infty. \tag{2.13}$$

- Advantage: the signal processing is very simple since the BS just multiplies the received vector with the conjugate-transpose of the channel matrix $\boldsymbol{H}$, and then detects each stream separately. More importantly, MRC can be implemented in a distributed manner. Furthermore, at low p_{u}, $\mathtt{SINR}_{\mathtt{mrc},k} \approx p_{\mathrm{u}}\|\boldsymbol{h}_k\|^2$. This implies that at low SNR, MRC can achieve the same array gain as in the case of a single-user system.
- Disadvantage: as discussed above, since MRC neglects the effect of multiuser interference, it performs poorly in interference-limited scenarios. This can be seen in (2.13), where the SINR is upper bounded by a constant (with respect to p_{u}) when p_{u} is large.

b) *Zero-Forcing Receiver:*

By contrast to MRC, zero-forcing (ZF) receivers take the interuser interference into account, but neglect the effect of noise. With ZF, the multiuser interference is completely nulled out by projecting each stream onto the orthogonal complement of the interuser interference. More precisely, the kth column of the ZF receiver matrix satisfies:

$$\begin{cases} \boldsymbol{a}_{\mathtt{zf},k}^H \boldsymbol{h}_k \neq 0 \\ \boldsymbol{a}_{\mathtt{zf},k}^H \boldsymbol{h}_{k'} = 0, \quad \forall k' \neq k. \end{cases} \tag{2.14}$$

The ZF receiver matrix, which satisfies (2.14) for all k, is the pseudo-inverse of the channel matrix $\boldsymbol{H}$. With ZF, we have

$$\tilde{\boldsymbol{y}}_{\mathrm{ul}} = \left(\boldsymbol{H}^H \boldsymbol{H}\right)^{-1} \boldsymbol{H}^H \boldsymbol{y}_{\mathrm{ul}} = \sqrt{p_{\mathrm{u}}}\boldsymbol{s} + \left(\boldsymbol{H}^H \boldsymbol{H}\right)^{-1} \boldsymbol{H}^H \boldsymbol{n}. \tag{2.15}$$

This scheme requires that $M \geq K$ (so that the matrix $\boldsymbol{H}^H\boldsymbol{H}$ is invertible). We can see that each stream (element) of $\tilde{\boldsymbol{y}}_{\text{ul}}$ in (2.15) is free of multiuser interference. The kth stream of $\tilde{\boldsymbol{y}}_{\text{ul}}$ is used to detect s_k:

$$\tilde{y}_{\text{ul},k} = \sqrt{p_{\text{u}}} s_k + \tilde{n}_k, \tag{2.16}$$

where $\tilde{n}_k$ denotes the kth element of $\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H\boldsymbol{n}$. Thus, the received SINR of the kth stream is given by

$$\texttt{SINR}_{\texttt{zf},k} = \frac{p_{\text{u}}}{\left[\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\right]_{kk}}. \tag{2.17}$$

- Advantage: the signal processing is simple and ZF works well in interference-limited scenarios. The SINR can be made as high as desired by increasing the transmit power.
- Disadvantage: since ZF neglects the effect of noise, it works poorly under noise-limited scenarios. Furthermore, if the channel is not well-conditioned then the pseudo-inverse amplifies the noise significantly, and hence, the performance is very poor. Compared with MRC, ZF has a higher implementation complexity due to the computation of the pseudo-inverse of the channel gain matrix.

c) *Minimum Mean-Square Error Receiver:*

The linear minimum mean-square error (MMSE) receiver aims to minimize the mean-square error between the estimate $\boldsymbol{A}^H\boldsymbol{y}_{\text{ul}}$ and the transmitted signal $\boldsymbol{s}$. More precisely,

$$\boldsymbol{A}_{\texttt{mmse}} = \underset{\boldsymbol{A}\in\mathbb{C}^{M\times K}}{\arg\min}\, \mathbb{E}\left\{\left\|\boldsymbol{A}^H\boldsymbol{y}_{\text{ul}} - \boldsymbol{s}\right\|^2\right\} \tag{2.18}$$

$$= \underset{\boldsymbol{A}\in\mathbb{C}^{M\times K}}{\arg\min}\, \sum_{k=1}^{K} \mathbb{E}\left\{\left|\boldsymbol{a}_k^H\boldsymbol{y}_{\text{ul}} - s_k\right|^2\right\}. \tag{2.19}$$

where $\boldsymbol{a}_k$ is the kth column of $\boldsymbol{A}$. Therefore, the kth column of the MMSE receiver matrix is [47]

$$\boldsymbol{a}_{\texttt{mmse},k} = \underset{\boldsymbol{a}_k\in\mathbb{C}^{M\times 1}}{\arg\min}\, \mathbb{E}\left\{\left|\boldsymbol{a}_k^H\boldsymbol{y}_{\text{ul}} - s_k\right|^2\right\} \tag{2.20}$$

$$= \text{cov}\left(\boldsymbol{y}_{\text{ul}}, \boldsymbol{y}_{\text{ul}}\right)^{-1} \text{cov}\left(s_k, \boldsymbol{y}_{\text{ul}}\right)^H \tag{2.21}$$

$$= \sqrt{p_{\text{u}}}\left(p_{\text{u}}\boldsymbol{H}\boldsymbol{H}^H + \boldsymbol{I}_M\right)^{-1}\boldsymbol{h}_k, \tag{2.22}$$

where $\text{cov}\left(\boldsymbol{v}_1, \boldsymbol{v}_2\right) \triangleq \mathbb{E}\left\{\boldsymbol{v}_1\boldsymbol{v}_2^H\right\}$, where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are two random column vectors with zero-mean elements.

It is known that the MMSE receiver maximizes the received SINR. Therefore, among the MMSE, ZF, and MRC receivers, MMSE is the best. We can see

from (2.22) that, at high SNR (high p_{u}), ZF approaches MMSE, while at low SNR, MRC performs as well as MMSE. Furthermore, substituting (2.22) into (2.10), the received SINR for the MMSE receiver is given by

$$\mathtt{SINR}_{\mathtt{mmse},k} = p_{\mathrm{u}} \boldsymbol{h}_k^H \left(p_{\mathrm{u}} \sum_{i \neq k}^{K} \boldsymbol{h}_i \boldsymbol{h}_i^H + \boldsymbol{I}_M \right)^{-1} \boldsymbol{h}_k. \tag{2.23}$$

2.4.2 Linear Precoders (in the Downlink)

In the downlink, with linear precoding techniques, the signal transmitted from M antennas, $\boldsymbol{x}$, is a linear combination of the symbols intended for the K users. Let q_k, $\mathbb{E}\left\{|q_k|^2\right\} = 1$, be the symbol intended for the kth user. Then, the linearly precoded signal vector $\boldsymbol{x}$ is

$$\boldsymbol{x} = \sqrt{\alpha} \boldsymbol{W} \boldsymbol{q}, \tag{2.24}$$

where $\boldsymbol{q} \triangleq [q_1 \ q_2 \ \ldots \ q_K]^T$, $\boldsymbol{W} \in \mathbb{C}^{M \times K}$ is the precoding matrix, and α is a normalization constant chosen to satisfy the power constraint $\mathbb{E}\left\{\|\boldsymbol{x}\|^2\right\} = 1$. Thus,

$$\alpha = \frac{1}{\mathbb{E}\left\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{W}^H)\right\}}. \tag{2.25}$$

A block diagram of the linear precoder at the BS is shown in Figure 2.3.

Plugging (2.24) into (2.4), we obtain

$$y_{\mathrm{dl},k} = \sqrt{\alpha p_{\mathrm{d}}} \boldsymbol{h}_k^T \boldsymbol{W} \boldsymbol{q} + z_k \tag{2.26}$$

$$= \sqrt{\alpha p_{\mathrm{d}}} \boldsymbol{h}_k^T \boldsymbol{w}_k q_k + \sqrt{\alpha p_{\mathrm{d}}} \sum_{k' \neq k}^{K} \boldsymbol{h}_k^T \boldsymbol{w}_{k'} q_{k'} + z_k. \tag{2.27}$$

Therefore, the SINR of the transmission from the BS to the kth user is

$$\mathtt{SINR}_k = \frac{\alpha p_{\mathrm{d}} \left| \boldsymbol{h}_k^T \boldsymbol{w}_k \right|^2}{\alpha p_{\mathrm{d}} \sum_{k' \neq k}^{K} \left| \boldsymbol{h}_k^T \boldsymbol{w}_{k'} \right|^2 + 1}. \tag{2.28}$$

Three conventional linear precoders are maximum-ratio transmission (MRT) (also called conjugate beforming), ZF, and MMSE precoders. These precoders have similar operational meanings and properties as MRC, ZF, MMSE receivers, respectively. Thus, here we just provide the final formulas for these precoders, i.e.,

$$\boldsymbol{W} = \begin{cases} \boldsymbol{H}^*, & \text{for MRT} \\ \boldsymbol{H}^* \left(\boldsymbol{H}^T \boldsymbol{H}^* \right)^{-1}, & \text{for ZF} \\ \boldsymbol{H}^* \left(\boldsymbol{H}^T \boldsymbol{H}^* + \frac{K}{p_{\mathrm{d}}} \boldsymbol{I}_K \right)^{-1}, & \text{for MMSE} \end{cases} \tag{2.29}$$

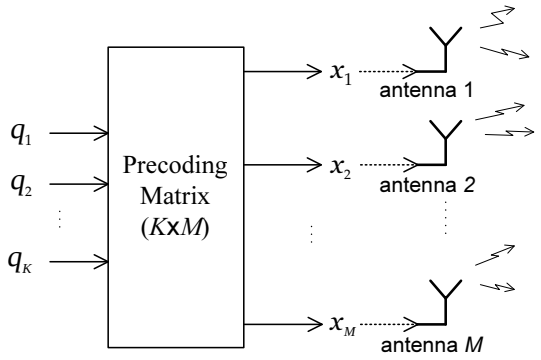


Figure 2.3: Block diagram of the linear precoders at the BS.

Figures 2.4 and 2.5 show the achievable sum rates for the uplink and the downlink transmission, respectively, with different linear processing schemes, versus $\mathsf{SNR} \triangleq p_{\mathrm{u}}$ for the uplink and $\mathsf{SNR} \triangleq p_{\mathrm{d}}$ for the downlink, with $M = 6$ and $K = 4$. The sum rate is defined as $\sum_{k=1}^{K} \mathbb{E}\left\{\log_2\left(1 + \mathtt{SINR}_k\right)\right\}$, where $\mathtt{SINR}_k$ is the SINR of the kth user which is given in the previous discussion. As expected, MMSE performs strictly better than ZF and MRC over the entire range of SNRs. In the low SNR regime, MRC is better than ZF, and vice versa in the high SNR regime.

2.5 Channel Estimation

We have assumed so far that the BS and the users have perfect CSI. However, in practice, this CSI has to be estimated. Depending on the system duplexing mode (TDD or FDD), the channel estimation schemes are very different.

2.5.1 Channel Estimation in TDD Systems

In a TDD system, the uplink and downlink transmissions use the same frequency spectrum, but different time slots. The uplink and downlink channels are reciprocal.[1] Thus, the CSI can be obtained by using following scheme (see Figure 2.6):

- For the uplink transmission: the BS needs CSI to detect the signals transmitted from the K users. This CSI is estimated at the BS. More precisely, the K users send K orthogonal pilot sequences to the BS on the uplink. Then the BS estimates the channels based on the received pilot signals. This process requires a minimum of K channel uses.

[1] In practice, the uplink and downlink channels are not perfectly reciprocal due to mismatches of the hardware chains. This non-reciprocity can be removed by calibration [15, 29, 30]. In our work, we assume that the hardware chain calibration is perfect.

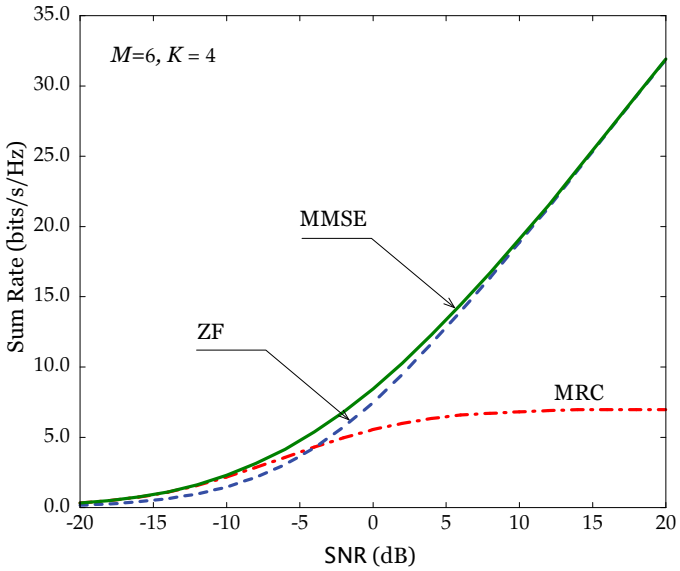


Figure 2.4: Performance of linear receivers in the uplink.

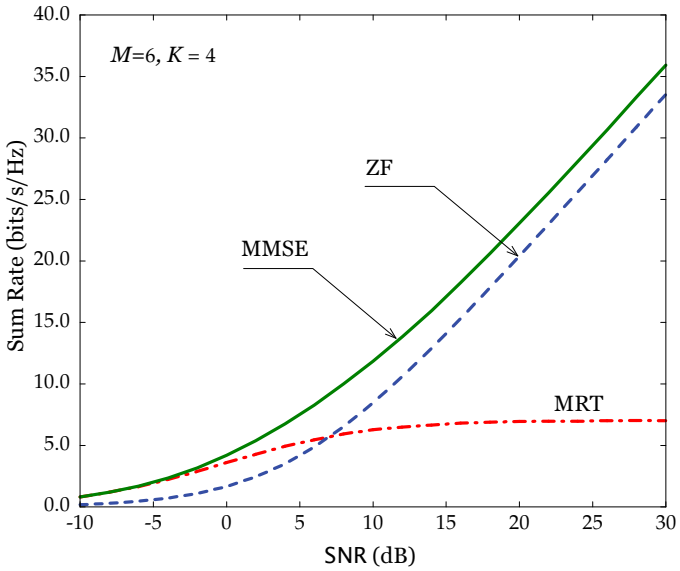


Figure 2.5: Performance of linear precoders in the downlink.

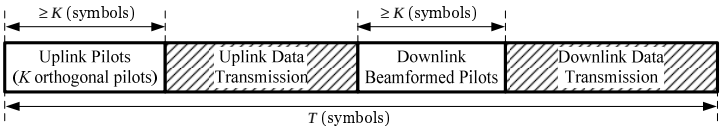


Figure 2.6: Slot structure and channel estimation in TDD systems.

- For the downlink: the BS needs CSI to precode the transmitted signals, while each user needs the effective channel gain to detect the desired signals. Due to the channel reciprocity, the channel estimated at the BS in the uplink can be used to precode the transmit symbols. To obtain knowledge of the effective channel gain, the BS can beamform pilots, and each user can estimate the effective channel gains based on the received pilot signals. This requires at least K channel uses.[2]

In total, the training process requires a minimum of $2K$ channel uses. We assume that the channel stays constant over T symbols. Thus, it is required that $2K < T$. An illustration of channel estimation in TDD systems is shown in Figure 2.6.

2.5.2 Channel Estimation in FDD Systems

In an FDD system, the uplink and downlink transmissions use different frequency spectrum, and hence, the uplink and downlink channels are not reciprocal. The channel knowledge at the BS and users can be obtained by using following training scheme:

- For the downlink transmission: the BS needs CSI to precode the symbols before transmitting to the K users. The M BS antennas transmit M orthogonal pilot sequences to K users. Each user will estimate the channel based on the received pilots. Then it feeds back its channel estimates (M channel estimates) to the BS through the uplink. This process requires at least M channel uses for the downlink and M channel uses for the uplink.
- For the uplink transmission: the BS needs CSI to decode the signals transmitted from the K users. One simple way is that the K users transmit K orthogonal pilot sequences to the BS. Then, the BS will estimate the channels based on the received pilot signals. This process requires at least K channel uses for the uplink.

[2]The effective channel gains at the users may be blindly estimated based on the received data, and hence, no pilots are required [31]. But, we do not discuss in detail about this possibility in this section.

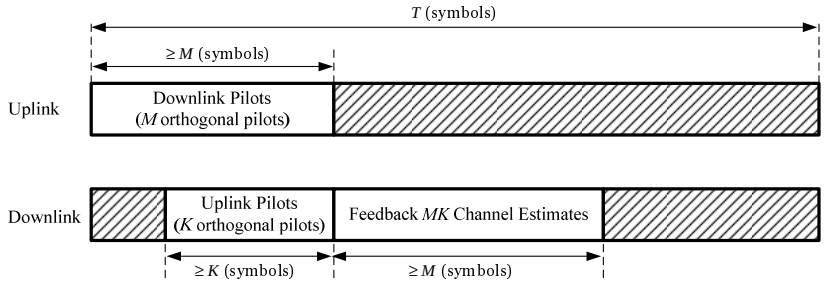


Figure 2.7: Slot structure and channel estimation in FDD systems.

Therefore, the entire channel estimation process requires at least $M + K$ channel uses in the uplink and M channel uses in the downlink. Assume that the lengths of the coherence intervals for the uplink and the downlink are the same and are equal to T. Then we have the constraints: $M < T$ and $M + K < T$. As a result $M+K < T$ is the constraint for FDD systems. An illustration of channel estimation in FDD systems is shown in Figure 2.7.

Chapter 3

Massive MIMO

3.1 What is Massive MIMO?

Massive MIMO is a form of MU-MIMO systems where the number of BS antennas and the numbers of users are large. In Massive MIMO, hundreds or thousands of BS antennas simultaneously serve tens or hundreds of users in the same frequency resource. Some main points of Massive MIMO are:

- TDD operation: as discussed in Section 2.5, with FDD, the channel estimation overhead depends on the number of BS antennas, M. By contrast, with TDD, the channel estimation overhead is independent of M. In Massive MIMO, M is large, and hence, TDD operation is preferable. For example, assume that the coherence interval is $T = 200$ symbols (corresponding to a coherence bandwidth of 200 kHz and a coherence time of 1 ms). Then, in FDD systems, the number of BS antennas and the number of users are constrained by $M + K < 200$, while in TDD systems, the constraint on M and K is $2K < 200$. Figure 3.1 shows the regions of feasible (M, K) in FDD and TDD systems. We can see that the FDD region is much smaller than the TDD region. With TDD, adding more antennas does not affect the resources needed for the channel estimation.
- Linear processing: since the number of BS antennas and the number of users are large, the signal processing at the terminal ends must deal with large dimensional matrices/vectors. Thus, simple signal processing is preferable. In Massive MIMO, linear processing (linear combing schemes in the uplink and linear precoding schemes in the downlink) is nearly optimal.

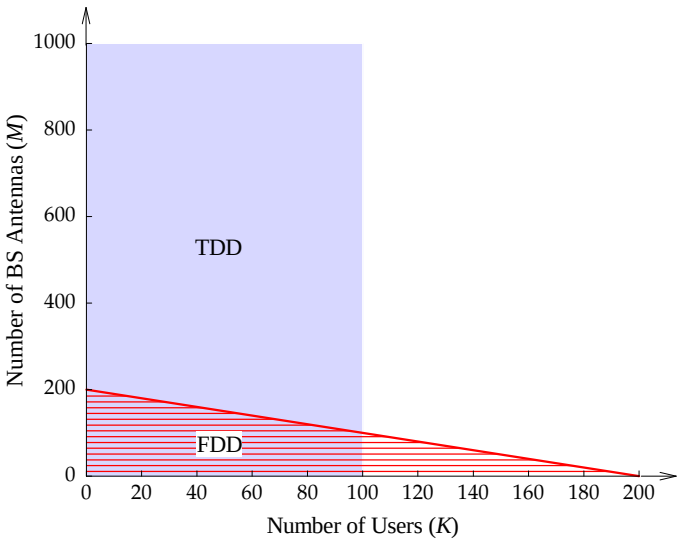


Figure 3.1: The regions of possible (M, K) in TDD and FDD systems, for a coherence interval of 200 symbols.

- Favorable propagation: favorable propagation means that the channel matrix between the BS antenna array and the users is well-conditioned. In Massive MIMO, under some conditions, the favorable propagation property holds due to the law of large numbers.
- A massive BS antenna array does not have to be physically large. For example consider a cylindrical array with 128 antennas, comprising four circles of 16 dual-polarized antenna elements. At 2.6 GHz, the distance between adjacent antennas is about 6 cm, which is half a wavelength, and hence, this array occupies only a physical size of 28cm×29cm [25].
- Massive MIMO is scalable: in Massive MIMO, the BS learns the channels via uplink training, under TDD operation. The time required for channel estimation is independent of the number of BS antennas. Therefore, the number of BS antennas can be made as large as desired with no increase in the channel estimation overhead. Furthermore, the signal processing at each user is very simple and does not depend on other users' existence, i.e., no multiplexing or de-multiplexing signal processing is performed at the users. Adding or dropping some users from service does not affect other users' activities.
- All the complexity is at the BS.

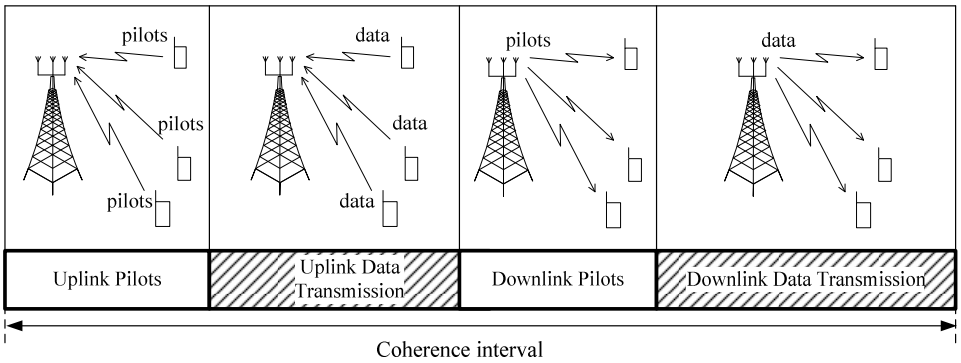


Figure 3.2: Transmission protocol of TDD Massive MIMO.

3.2 How Massive MIMO Works

In Massive MIMO, TDD operation is preferable. During a coherence interval, there are three operations: channel estimation (including the uplink training and the downlink training), uplink data transmission, and downlink data transmission. A TDD Massive MIMO protocol is shown in Figure 3.2.

3.2.1 Channel Estimation

The BS needs CSI to detect the signals transmitted from the users in the uplink, and to precode the signals in the downlink. This CSI is obtained through the uplink training. Each user is assigned an orthogonal pilot sequence, and sends this pilot sequence to the BS. The BS knows the pilots sequences transmitted from all users, and then estimates the channels based the received pilot signals. The estimation schemes were discussed in detail in Section 2.5.1.

Furthermore, each user may need partial knowledge of CSI to coherently detect the signals transmitted from the BS. This information can be acquired through downlink training or some blind channel estimation algorithm. Since the BS uses linear precoding techniques to beamform the signals to the users, the user needs only the effective channel gain (which is a scalar constant) to detect its desired signals. Therefore, the BS can spend a short time to beamform pilots in the downlink for CSI acquisition at the users.

3.2.2 Uplink Data Transmission

A part of the coherence interval is used for the uplink data transmission. In the uplink, all K users transmit their data to the BS in the same time-frequency resource. The BS then uses the channel estimates together with the linear combining

techniques to detect signals transmitted from all users. The detailed uplink data transmission was discussed in Section 2.2.

3.2.3 Downlink Data Transmission

In the downlink, the BS transmits signals to all K users in the same time-frequency resource. More specifically, the BS uses its channel estimates in combination with the symbols intended for the K users to create M precoded signals which are then fed to M antennas. The downlink data transmission was discussed in detail in Section 2.3.

3.3 Why Massive MIMO

The demand for wireless throughput and communication reliability as well as the user density will always increase. Future wireless commination requires new technologies in which many users can be simultaneously served with very high throughput. Massive MIMO can meet these demands. Consider the uplink transmission. (The same argument can be used for the downlink transmission.) From (2.3), under the conditions of favorable propagation (the channel vectors between the users and the BS are pairwisely orthogonal), the sum-capacity of the uplink transmission is

$$C_{\text{sum}} = \log_2 \det \left(\boldsymbol{I}_K + p_{\text{u}} M \boldsymbol{I}_K \right) = K \log_2 (1 + M p_{\text{u}}). \tag{3.1}$$

In (3.1), K is the multiplexing gain, and M represents the array gain. We can see that, we can obtain a huge spectral efficiency and energy efficiency when M and K are large. Without any increase in transmitted power per terminal, by increasing K and M, we can simultaneously serve more users in the same frequency band. At the same time the throughput per user also increases. Furthermore, by doubling the number of BS antennas, we can reduce the transmit power by 3 dB, while maintaining the original quality-of-service.

The above gains (multiplexing gain and array gain) are obtained under the conditions of favorable propagation and the use of optimal processing at the BS. One main question is: Will these gains still be obtained by using linear processing? Another question is: Why not use the conventional low dimensional point-to-point MIMO with complicated processing schemes instead of Massive MIMO with simple linear processing schemes? In Massive MIMO, when the number of BS antennas is large, due to the law of large numbers, the channels become favorable. As a result, linear processing is nearly optimal. The multiplexing gain and array gain can be obtained with simple linear processing. Also, by increasing the number of BS antennas and the number of users, we can always increase the throughput.

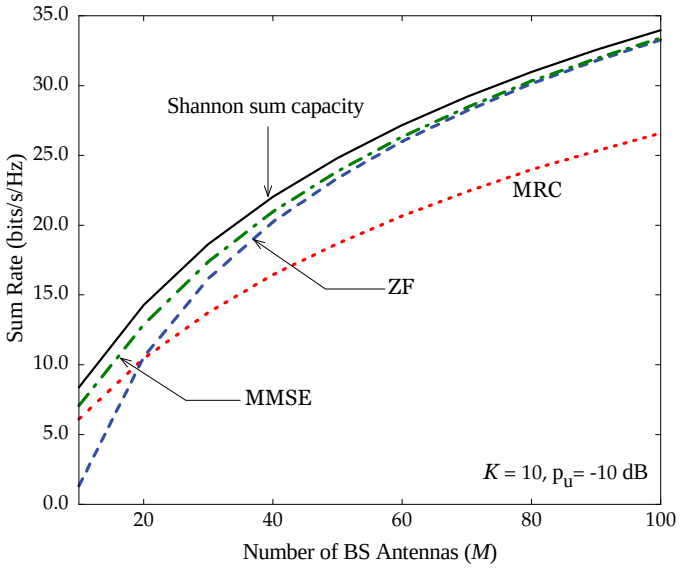


Figure 3.3: Uplink sum rate for different linear receivers and for the optimal receiver.

Figure 3.3 shows the sum rate versus the number of BS antennas with optimal receivers (the sum capacity is achieved) and linear receivers, at $K = 10$ and $p_{\mathrm{u}} = -10$ dB. The sum capacity is computed from (2.3), while the sum rates for MRC, ZF, and MMSE are computed by using (2.12), (2.17), and (2.23), respectively. We can see that, when M is large, the sum rate with linear processing is very close to the sum capacity obtained by using optimal receivers. When $M = K = 10$, and with the optimal receiver, the maximum sum rate that we can obtain is 8.5 bits/s/Hz. By contrast, by using large M, say $M = 50$, with simple ZF receivers, we can obtain a sum rate of 24 bits/s/HZ.

3.4 Challenges in Massive MIMO

Despite the huge advantages of Massive MIMO, many issues still need to be tackled. The main challenges of Massive MIMO are listed as follows:

3.4.1 Pilot Contamination

In previous sections, we considered single-cell setups. However, practical cellular networks consist of many cells. Owing to the limited availability of frequency spectrum, many cells have to share the same time-frequency resources. Thus, multicell

setups should be considered. In multicell systems, we cannot assign orthogonal pilot sequences for all users in all cells, due to the limitation of the channel coherence interval. Orthogonal pilot sequences have to be reused from cell to cell. Therefore, the channel estimate obtained in a given cell will be contaminated by pilots transmitted by users in other cells. This effect, called "pilot contamination", reduces the system performance [32]. The effect of pilot contamination is major inherent limitation of Massive MIMO. It does not vanish even when the number of BS antennas grows without bound. Considerable efforts have been made to reduce this effect. The eigenvalue-decomposition-based channel estimation, pilot decontamination, as well as pilot contamination precoding schemes are proposed in [33–35]. In [36], the authors shown that, under certain conditions of the channel covariance, by using a covariance aware pilot assignment scheme among the cells, pilot contamination can be efficiently mitigated. There is much ongoing research on this topic.

3.4.2 Unfavorable Propagation

Massive MIMO works under favorable propagation environments. However, in practice, there may be propagation environments where the channels are not favorable. For example, in propagation environments where the numbers of the scatterers is small compared to the numbers of users, or the channels from different users to the BS share some common scatterers, the channel is not favorable [31]. One possibility to tackle this problem is to distribute the BS antennas over a large area.

3.4.3 New Standards and Designs are Required

It will be very efficient if Massive MIMO can be deployed in current systems such as LTE. However, the LTE standard only allows for up to 8 antenna ports at the BS [4]. Furthermore, LTE uses the channel information that is "assumed". For example, one option of the downlink in LTE is that the BS transmits the reference signals through several fixed beams. Then the users report back to the BS the strongest beam. The BS will use this beam for the downlink transmission. By contrast, Massive MIMO uses the channel information that is estimated (measured). Therefore, to reduce Massive MIMO to practice, new standards are required. On a different note, with Massive MIMO, a costly 40 Watt transceiver should be replaced by a large number of low-power and inexpensive antennas. Related hardware designs should also be considered. This requires a huge effort from both academia and industry.

Chapter 4

Mathematical Preliminaries

4.1 Random Matrix Theory

We now review some useful limit results about very long random vectors [37] which will be used for the analysis in the rest of the thesis.

- Let $\boldsymbol{p} \triangleq [p_1 \ ... \ p_n]^T$ and $\boldsymbol{q} \triangleq [q_1 \ ... \ q_n]^T$ be $n \times 1$ vectors whose elements are independent identically distributed (i.i.d.) random variables (RVs) with $\mathbb{E}\{p_i\} = \mathbb{E}\{q_i\} = 0$, $\mathbb{E}\left\{|p_i|^2\right\} = \sigma_p^2$, and $\mathbb{E}\left\{|q_i|^2\right\} = \sigma_q^2$, $i = 1, 2, ..., n$. Assume that $\boldsymbol{p}$ and $\boldsymbol{q}$ are independent.

  Applying the law of large numbers, we obtain

$$\frac{1}{n}\boldsymbol{p}^H\boldsymbol{p} \overset{a.s.}{\to} \sigma_p^2, \text{ as } n \to \infty, \tag{4.1}$$

$$\frac{1}{n}\boldsymbol{p}^H\boldsymbol{q} \overset{a.s.}{\to} 0, \text{ as } n \to \infty, \tag{4.2}$$

  where $\overset{a.s.}{\to}$ denotes almost sure convergence.

  Applying the Lindeberg-Lévy central limit theorem, we obtain

$$\frac{1}{\sqrt{n}}\boldsymbol{p}^H\boldsymbol{q} \overset{d}{\to} \mathcal{CN}\left(0, \sigma_p^2\sigma_q^2\right), \text{ as } n \to \infty, \tag{4.3}$$

  where $\overset{d}{\to}$ denotes convergence in distribution.

- Let X_1, X_2, ... be a sequence of independent circularly symmetric complex RVs, such that X_i has zero mean and variance σ_i^2. Further assume that the following conditions are satisfied: 1) $s_n^2 = \sum_{i=1}^{n} \sigma_i^2 \to \infty$, as $n \to \infty$; and 2) $\sigma_i / s_n \to 0$, as $n \to \infty$. Then by applying the Cramér's central limit theorem, we have

$$\frac{\sum_{i=1}^{n} X_i}{s_n} \xrightarrow{d} \mathcal{CN}(0,1), \text{ as } n \to \infty. \tag{4.4}$$

- Let $X_1, X_2, ...X_n$ be independent RVs such that $\mathbb{E}\{X_i\} = \mu_i$ and $\mathbb{V}\text{ar}(X_i) < c < \infty$, $\forall i = 1, \ldots, n$. Then by applying the Tchebyshev's theorem, we have

$$\frac{1}{n}(X_1 + X_2 + ... + X_n) - \frac{1}{n}(\mu_1 + \mu_2 + ...\mu_n) \xrightarrow{P} 0, \tag{4.5}$$

  where $\xrightarrow{P}$ denotes convergence in probability.

4.2 Capacity Lower Bounds

In this section, we derive a capacity lower bound for a SISO channel with interference and with partial/perfect CSI at the receivers. The channel model is:

$$y_k = \underbrace{a_k s_k}_{\text{desired signal}} + \underbrace{\sum_{k' \neq k}^{K} a_{k'} s_{k'}}_{\text{interference}} + \underbrace{n_k}_{\text{noise}}, \tag{4.6}$$

where a_k, $k = 1, ..., K$, are the effective channel gains, s_k, is the transmitted signals from the kth sources, and n_k is the additive noise. Assume that s_k, $k = 1, ..., K$, and n_k are independent RVs with zero-mean and unit variance.

Since in this thesis, we consider linear processing, the end-to-end channel can be considered as an interference SISO channel, and can be modeled as in (4.6). For example, consider the downlink transmission discussed in Section 2.4. The received signal at the kth user is given by (2.27) which precisely matches with the model (4.6), where $a_{k'}$ is $\sqrt{\alpha p_{\text{d}}} \boldsymbol{h}_k^T \boldsymbol{w}_{k'}$, $k' = 1, ..., K$. The capacity lower bound derived in this section will be used throughout the thesis.

Let $\mathcal{C}$ be the channel state information (CSI) available at the receiver. We assume that $s_k \sim \mathcal{CN}(0,1)$. In general, Gaussian signaling is not optimal, and hence, this assumption yields a lower bound on the capacity:

$$C_k = I(s_k; y_k, \mathcal{C}) = h(s_k) - h(s_k | y_k, \mathcal{C}) \tag{4.7}$$

$$\overset{(a)}{=} \log_2(\pi e) - h(s_k - \alpha y_k | y_k, \mathcal{C}) \tag{4.8}$$

$$\overset{(b)}{\geq} \log_2(\pi e) - h(s_k - \alpha y_k | \mathcal{C}), \tag{4.9}$$

where (a) holds for any α, and (b) follows from the fact that conditioning reduces the entropy. Note that, in (4.7), $I(x;y)$ and $h(x)$ denote the mutual information between x and y, and the differential entropy of x, respectively.

Since the differential entropy of a RV with fixed variance is maximized when the RV is Gaussian, we obtain

$$C_k \geq \log_2(\pi e) - \mathbb{E}\left\{\log_2\left(\pi e\,\mathbb{E}\left\{\left|s_k - \alpha y_k\right|^2 \middle| \mathcal{C}\right\}\right)\right\}, \tag{4.10}$$

which leads to

$$C_k \geq \mathbb{E}\left\{\log_2\left(\frac{1}{\mathbb{E}\left\{\left|s_k - \alpha y_k\right|^2 \middle| \mathcal{C}\right\}}\right)\right\}. \tag{4.11}$$

To obtain the tightest bound, we choose $\alpha = \alpha_0$ so that $\mathbb{E}\left\{\left|x_k - \alpha_0 y_k\right|^2 \middle| \mathcal{C}\right\}$ is minimized:

$$\alpha_0 = \arg\min_{\alpha} \mathbb{E}\left\{\left|s_k - \alpha y_k\right|^2 \middle| \mathcal{C}\right\}. \tag{4.12}$$

We can see that $\alpha_0 y_k$ is the LMMSE estimate of s_k. Therefore,

$$\alpha_0 = \frac{\mathbb{E}\left\{y_k^* s_k | \mathcal{C}\right\}}{\mathbb{E}\left\{\left|y_k\right|^2 \middle| \mathcal{C}\right\}} = \frac{\mathbb{E}\left\{a_k^* | \mathcal{C}\right\}}{\mathbb{E}\left\{\left|y_k\right|^2 \middle| \mathcal{C}\right\}}. \tag{4.13}$$

We have

$$\begin{aligned}\mathbb{E}\left\{\left|s_k - \alpha y_k\right|^2 \middle| \mathcal{C}\right\} &= \mathbb{E}\left\{\left|s_k\right|^2\right\} - \alpha^* \mathbb{E}\left\{s_k y_k^* | \mathcal{C}\right\} - \alpha\,\mathbb{E}\left\{s_k^* y_k | \mathcal{C}\right\} + \left|\alpha\right|^2 \mathbb{E}\left\{\left|y_k\right|^2 \middle| \mathcal{C}\right\} \\ &= 1 - \alpha^* \mathbb{E}\left\{a_k^* | \mathcal{C}\right\} - \alpha\,\mathbb{E}\left\{a_k | \mathcal{C}\right\} + \left|\alpha\right|^2 \mathbb{E}\left\{\left|y_k\right|^2 \middle| \mathcal{C}\right\}. \end{aligned} \tag{4.14}$$

When $\alpha = \alpha_0$, substituting (4.13) into (4.14), we get

$$\mathbb{E}\left\{\left|x_k - \alpha y_k\right|^2 \middle| \mathcal{C}\right\} = 1 - \frac{\left|\mathbb{E}\left\{a_k^* | \mathcal{C}\right\}\right|^2}{\sum_{k=1}^{K} \mathbb{E}\left\{\left|a_k\right|^2 \middle| \mathcal{C}\right\} + 1}. \tag{4.15}$$

Plugging (4.15) into (11), we obtain the following lower bound on the capacity of (4.6):

$$C_k \geq \mathbb{E}\left\{\log_2\left(1 + \frac{\left|\mathbb{E}\left\{a_k | \mathcal{C}\right\}\right|^2}{\sum_{k'=1}^{K} \mathbb{E}\left\{\left|a_{k'}\right|^2 \middle| \mathcal{C}\right\} - \left|\mathbb{E}\left\{a_k | \mathcal{C}\right\}\right|^2 + 1}\right)\right\}. \tag{4.16}$$

We next consider two special cases:

1. No instantaneous CSI: for this case, $\mathcal{C} = \emptyset$. Therefore, from (2), we obtain

$$C_k \geq \log_2 \left(1 + \frac{\left| \mathbb{E}\left\{ a_k \right\} \right|^2}{\mathbb{E}\left\{ \left| a_k - \mathbb{E}\left\{ a_k \right\} \right|^2 \right\} + \sum_{k' \neq k}^{K} \mathbb{E}\left\{ \left| a_{k'} \right|^2 \right\} + 1} \right). \tag{4.17}$$

   This bound is often used in Massive MIMO research since it has a simple closed-form solution. Furthermore, in most propagation environments, when the number of BS antennas is large, the channel hardens (the effective channel gains become deterministic), and hence, this bound is very tight.

2. Full CSI: in this case, $\mathcal{C} = \{a_1, \ldots, a_K\}$. Substituting $\mathcal{C} = \{a_1, \ldots, a_K\}$ into (2), we obtain

$$C_k \geq \mathbb{E}\left\{ \log_2 \left(1 + \frac{\left| a_k \right|^2}{\sum_{k' \neq k}^{K} \left| a_{k'} \right|^2 + 1} \right) \right\}. \tag{4.18}$$

Chapter 5

Summary of Specific Contributions of the Dissertation

This dissertation consists of two parts. Firstly, we study fundamentals of Massive MIMO. The performance of Massive MIMO systems is analysed in terms of spectral efficiency and energy efficiency. Effects of pilot contamination and finite-dimensional channel models are also analysed. In addition, some aspects of favorable propagation in Massive MIMO are investigated. Secondly, we propose some system designs for Massive MIMO. Specifically, the optimal power as well as training duration allocation is studied. Applications of Massive MIMO in relay channels are also considered.

5.1 Included Papers

Brief summaries of the papers included in this dissertation are as follows:

Paper A: Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems

Authored by Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta.

Published in the IEEE Transactions on Communications, 2013. This work is an extension of the conference paper [38].

A multiplicity of autonomous terminals simultaneously transmits data streams to a compact array of antennas. The array uses imperfect channel-state information derived from transmitted pilots to extract the individual data streams. The power radiated by the terminals can be made inversely proportional to the square-root of the number of base station antennas with no reduction in performance. In contrast if perfect channel-state information were available the power could be made inversely proportional to the number of antennas. Lower capacity bounds for maximum-ratio combining (MRC), zero-forcing (ZF) and minimum mean-square error (MMSE) detection are derived. A MRC receiver normally performs worse than ZF and MMSE. However as power levels are reduced, the cross-talk introduced by the inferior maximum-ratio receiver eventually falls below the noise level and this simple receiver becomes a viable option. The tradeoff between the energy efficiency (as measured in bits/J) and spectral efficiency (as measured in bits/channel use/terminal) is quantified for a channel model that includes small-scale fading but not large-scale fading. It is shown that the use of moderately large antenna arrays can improve the spectral and energy efficiency with orders of magnitude compared to a single-antenna system.

Paper B: The Multicell Multiuser MIMO Uplink with Very Large Antenna Arrays and a Finite-Dimensional Channel

Authored by Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta.

Published in the IEEE Transactions on Communications, 2013. This work is an extension of the conference paper [39].

We consider multicell multiuser MIMO systems with a very large number of antennas at the base station (BS). We assume that the channel is estimated by using uplink training. We further consider a physical channel model where the angular domain is separated into a finite number of distinct directions. We analyze the so-called pilot contamination effect discovered in previous work, and show that this effect persists under the finite-dimensional channel model that we consider. In particular, we consider a uniform array at the BS. For this scenario, we show that when the number of BS antennas goes to infinity, the system performance under a finite-dimensional channel model with P angular bins is the same as the performance under an uncorrelated channel model with P antennas. We further derive a lower bound on the achievable rate of uplink data transmission with a linear detector at the BS. We then specialize this lower bound to the cases of maximum-ratio combining (MRC) and zero-forcing (ZF) receivers, for a finite and an infinite number of BS antennas. Numerical results corroborate our analysis and show a comparison between the performances of MRC and ZF in terms of sum-rate.

Paper C: Aspects of Favorable Propagation in Massive MIMO

Authored by Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta.

Published in the proceedings of the European Signal Processing Conference (EUSIPCO), 2014 [40].

Favorable propagation, defined as mutual orthogonality among the vector-valued channels to the terminals, is one of the key properties of the radio channel that is exploited in Massive MIMO. However, there has been little work that studies this topic in detail. In this paper, we first show that favorable propagation offers the most desirable scenario in terms of maximizing the sum-capacity. One useful proxy for whether propagation is favorable or not is the channel condition number. However, this proxy is not good for the case where the norms of the channel vectors are not equal. For this case, to evaluate how favorable the propagation offered by the channel is, we propose a "distance from favorable propagation" measure, which is the gap between the sum-capacity and the maximum capacity obtained under favorable propagation. Secondly, we examine how favorable the channels can be for two extreme scenarios: i.i.d. Rayleigh fading and uniform random line-of-sight (UR-LoS). Both environments offer (nearly) favorable propagation. Furthermore, to analyze the UR-LoS model, we propose an urns-and-balls model. This model is simple and explains the singular value spread characteristic of the UR-LoS model well.

Paper D: EVD-Based Channel Estimations for Multicell Multiuser MIMO with Very Large Antenna Arrays

Authored by Hien Quoc Ngo and Erik G. Larsson.

Published in the proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012 [33].

This paper consider a multicell multiuser MIMO with very large antenna arrays at the base station. For this system, with channel state information estimated from pilots, the system performance is limited by pilot contamination and noise limitation as well as the spectral inefficiency discovered in previous work. To reduce these effects, we propose the eigenvalue-decomposition-based approach to estimate the channel directly from the received data. This approach is based on the orthogonality of the channel vectors between the users and the base station when the number of base station antennas grows large. We show that the channel can be estimated from the eigenvalue of the received covariance matrix excepting the multiplicative factor ambiguity. A short training sequence is required to solved this ambiguity. Furthermore, to improve the performance of our approach, we investigate the join eigenvalue-decomposition-based approach and the Iterative Least-Square with Projection algorithm. The numerical results verify the effectiveness of our channel estimate approach.

Paper E: Massive MU-MIMO Downlink TDD Systems with Linear Precoding and Downlink Pilots

Authored by Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta.

Published in the proceedings of the Allerton Conference on Communication, Control, and Computing, 2013 [41].

We consider a massive MU-MIMO downlink time-division duplex system where a base station (BS) equipped with many antennas serves several single-antenna users in the same time-frequency resource. We assume that the BS uses linear precoding for the transmission. To reliably decode the signals transmitted from the BS, each user should have an estimate of its channel. In this work, we consider an efficient channel estimation scheme to acquire CSI at each user, called beamforming training scheme. With the beamforming training scheme, the BS precodes the pilot sequences and forwards to all users. Then, based on the received pilots, each user uses minimum mean-square error channel estimation to estimate the effective channel gains. The channel estimation overhead of this scheme does not depend on the number of BS antennas, and is only proportional to the number of users. We then derive a lower bound on the capacity for maximum-ratio transmission and zero-forcing precoding techniques which enables us to evaluate the spectral efficiency taking into account the spectral efficiency loss associated with the transmission of the downlink pilots. Comparing with previous work where each user uses only the statistical channel properties to decode the transmitted signals, we see that the proposed beamforming training scheme is preferable for moderate and low-mobility environments.

Paper F: Blind Estimation of Effective Downlink Channel Gains in Massive MIMO

Authored by Hien Quoc Ngo and Erik G. Larsson.

Submitted to the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015 [31].

We consider the massive MIMO downlink with time-division duplex (TDD) operation and conjugate beamforming transmission. To reliably decode the desired signals, the users need to know the effective channel gain. In this paper, we propose a blind channel estimation method which can be applied at the users and which does not require any downlink pilots. We show that our proposed scheme can substantially outperform the case where each user has only statistical channel knowledge, and that the difference in performance is particularly large in certain types of channel, most notably keyhole channels. Compared to schemes that rely on downlink pilots (e.g., [41]), our proposed scheme yields more accurate channel estimates for a wide range of signal-to-noise ratios and avoid spending time-frequency resources on pilots.

Paper G: Massive MIMO with Optimal Power and Training Duration Allocation

Authored by Hien Quoc Ngo, Michail Matthaiou, and Erik G. Larsson.

Published in the IEEE Wireless Communications Letters, 2014 [42].

We consider the uplink of massive multicell multiple-input multiple-output systems, where the base stations (BSs), equipped with massive arrays, serve simultaneously several terminals in the same frequency band. We assume that the BS estimates the channel from uplink training, and then uses the maximum ratio combining technique to detect the signals transmitted from all terminals in its own cell. We propose an optimal resource allocation scheme which jointly selects the training duration, training signal power, and data signal power in order to maximize the sum spectral efficiency, for a given total energy budget spent in a coherence interval. Numerical results verify the benefits of the optimal resource allocation scheme. Furthermore, we show that more training signal power should be used at low signal-to-noise ratio (SNRs), and vice versa at high SNRs. Interestingly, for the entire SNR regime, the optimal training duration is equal to the number of terminals.

Paper H: Large-Scale Multipair Two-Way Relay Networks with Distributed AF Beamforming

Authored by Hien Quoc Ngo and Erik G. Larsson.

Published in the IEEE Communications Letters, 2013 [43].

We consider a multipair two-way relay network where multiple communication pairs simultaneously exchange information with the help of multiple relay nodes. All nodes are equipped with a single antenna and channel state information is available at the relay nodes. Each relay uses very simple signal processing in a distributed manner, called distributed amplify-and-forward (AF) relaying. A closed-form expression for the achievable rate is derived. We show that the distributed AF scheme outperforms conventional orthogonal relaying. When the number of relays is large, the distributed AF relaying scheme can achieve the capacity scaling given by the cut-set upper bound. Furthermore, when the number of relays grows large, the transmit powers of each terminal and of the relay can be made inversely proportional to the number of relays while maintaining a given quality-of-service. If the transmit power of each terminal is kept fixed, the transmit power of each relay can be scaled down inversely proportional to the square of the number of relays.

Paper I: Spectral Efficiency of the Multi-pair Two-Way Relay Channel with Massive Arrays

Authored by Hien Quoc Ngo and Erik G. Larsson.

Published in the proceedings of the Asilomar Conference on Signals, Systems, and Computer, 2013 [44].

We consider a multipair two-way relay channel where multiple communication pairs share the same time-frequency resource and a common relay node. We assume that all users have a single antenna, while the relay node is equipped with a very large antenna array. We consider two transmission schemes: (I) separate-training zero-forcing (ZF) and (II) a new proposed coupled-training ZF. For both schemes, the channels are estimated at the relay by using training sequences, assuming time-division duplex operation. The relay processes the received signals using ZF. With the separate-training ZF, the channels from all users are estimated separately. By contrast, with the coupled-training ZF, the relay estimates the sum of the channels from two users of a given communication pair. This reduces the amount of resources spent in the training phase. Self-interference reduction is also proposed for these schemes. When the number of relay antennas grows large, the effects of interpair interference and self-interference can be neglected. The transmit power of each user and of the relay can be made inversely proportional to the square root of the number of relay antennas while maintaining a given quality-of-service. We derive a lower bound on the capacity which enables us to evaluate the spectral efficiency. The coupled-training ZF scheme is preferable for the high-mobility environment, while the separate-training ZF scheme is preferable for the low-mobility environment

Paper J: Multipair Full-Duplex Relaying with Massive Arrays and Linear Processing

Authored by Hien Quoc Ngo, Himal A. Suraweera, Michail Matthaiou, and Erik G. Larsson.

Published in the IEEE Journal on Selected Areas in Communications, 2014 [45].

We consider a multipair decode-and-forward relay channel, where multiple sources transmit simultaneously their signals to multiple destinations with the help of a full-duplex relay station. We assume that the relay station is equipped with massive arrays, while all sources and destinations have a single antenna. The relay station uses channel estimates obtained from received pilots and zero-forcing (ZF) or maximum-ratio combining/maximum-ratio transmission (MRC/MRT) to process the signals. To significantly reduce the loop interference effect, we propose two techniques: i) using a massive receive antenna array; or ii) using a massive transmit antenna array together with very low transmit power at the relay station. We derive an exact achievable rate expression in closed-form for MRC/MRT processing and an analytical approximation of the achievable rate for ZF processing. This approximation is very tight, particularly for a large number of relay station antennas. These closed-form expressions enable us to determine the regions where the full-duplex mode outperforms the half-duplex mode, as well as to design an optimal power allocation scheme. This optimal power allocation scheme aims to maximize the energy efficiency for a given sum spectral efficiency and under peak power constraints at the relay station and sources. Numerical results verify the effectiveness of the optimal power allocation scheme. Furthermore, we show that, by doubling the number of transmit/receive antennas at the relay station, the transmit

power of each source and of the relay station can be reduced by 1.5 dB if the pilot power is equal to the signal power, and by 3 dB if the pilot power is kept fixed, while maintaining a given quality of service.

5.2 Not Included Papers

The following publications by the author are not included in the dissertation either because they do not fit within the main scope of the dissertation, or they were earlier versions of the journal publications included in the dissertation.

- H. Q. Ngo and E. G. Larsson, "Linear multihop amplify-and-forward relay channels: Error exponent and optimal number of hops," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3834–3842, Nov. 2011.
- M. Matthaiou, G. C. Alexandropoulos, H. Q. Ngo, and E. G. Larsson, "Analytic framework for the effective rate of MISO fading channels," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1741–1751, June 2012.
- T. Q. Duong, H. Q. Ngo, H-J. Zepernick, and A. Nallanathan, "Distributed space-time coding in two-way fixed gain relay networks over Nakagami-m fading," *IEEE International Conference on Communicatons (ICC)*, Ottawa, Canada, June 2012.
- H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Performance analysis of large scale MU-MIMO with optimal linear receivers", *Proceedings of the IEEE Swedish Communication Technologies Workshop (Swe-CTW)*, 2012.
- H. Q. Ngo, M. Matthaiou, T. Q. Duong, and E. G. Larsson, "Uplink performance analysis of multicell MU-SIMO systems with ZF receivers," *IEEE Trans. Vehicular Techno.*, vol. 62, no. 9, pp. 4471–4483, Nov. 2013.
- H. A. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen, and E. G. Larsson, "Multi-pair amplify-and-forward relaying with very large antenna arrays," in *Proc. IEEE International Conference on Communicatons (ICC)*, Budapest, Hungary, June 2013.
- A. K. Papazafeiropoulos, H. Q. Ngo, M. Matthaiou, and T. Ratnarajah, "Uplink performance of conventional and massive MIMO cellular systems with delayed CSIT," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Washington, D.C., Sept. 2014.
- H. Phan, T. M. C. Chu, H.-J. Zepernick, H. Q. Ngo, "Performance of cognitive radio networks with finite buffer using multiple vacations and exhaustive service," in *Proc. International Conference on Signal Processing and Communication (ICSPCS)*, Gold Coast, Australia, Dec. 2014.

Chapter 6

Future Research Directions

As discussed in Section 3.4, a number of issues and challenges in Massive MIMO remain to be investigated. There are many open research directions. Here is the list of possible research directions in Massive MIMO:

- Pilot contamination: pilot contamination is one of the inherent limitations of Massive MIMO which degrades the Massive MIMO performance significantly. This effect persists even when the number of BS antennas goes to infinity. Dealing with the pilot contamination effect is an important research direction. Pilot contamination arises due to interference from other cells during the training phase. Therefore, one way to reduce the pilot contamination effect is to use large frequency-reuse factors during the training phase. However, this will reduce the pre-log factor, and hence, will reduce the spectral efficiency. Another way is increasing the cell-size. With a large cell-size, due to the path loss, the power of the desired signal in a given cell is much stronger the interference power from other cells. However, owing to the large cell-size and the effect of path loss, the users that are located around the cell edges could not receive a good quality-of-service. A suitable design of the cell-size, frequency-reuse factor during the training, and power control to reduce the pilot contamination effect should be investigated.
- Channel state information acquisition: the acquisition of CSI is very important in Massive MIMO. Channel estimation algorithms are attracting much attention. There is much ongoing research in this direction. Many questions are not still appropriately answered:
  - Can the channel be blindly estimated? Can payload data help improve channel estimation accuracy? How much can we gain from such schemes?

  - Is the use of orthogonal pilot sequences among users optimal, especially in multicell systems where the pilot contamination effect occurs? Which pilot sequences should be used? How to optimally assign pilot sequences to the users, especially for new users which enter the system? We believe that the design and assignment of the pilot sequences is an important research direction.
  - Should each user estimate the effective channel in the downlink? And how much gain we can obtain?

- System architecture: it would be good if Massive MIMO can combine with practical systems such as LTE. Furthermore, Massive MIMO, small cell, and millimeter wave technologies are promising candidates for 5G wireless systems. Designing new efficient systems with the combination of these technologies is a good research direction.

- In our work, Paper F, we considered the case where the transmit powers during the training and payload data transmissions are not equal, and are optimally chosen. The performance gain obtained by this optimal power allocation was studied. However, the cost of performing this optimal power allocation may be an increase in the peak-to-average ratio of the emitted waveform. This should be investigated in future work.

- In Papers G and H, amplify-and-forward relaying networks are studied. We assume that the destinations have no CSI, and an interpair interference reduction scheme is proposed by using the knowledge of its deterministic equivalent. This will work well if the interference hardens quickly in large systems. However, with amplifying-and-forward relaying, the interference contains the products of two channels, and hence, the interference hardens very slowly. The system performance may noticeably improve if the estimate of interference is considered.

- In current works, we assume that each user has a single antenna. It would be interesting to consider the case where each user is employed with several antennas. Note that in the current wireless systems (e.g. LTE), each users can have two antennas [48]. The transceiver designs (transmission schemes at the BS and detection schemes at the users) and performance analysis (achievable rate, outage probability, etc) of Massive MIMO systems with multiple-antenna users should be studied.

- Distributed Massive MIMO: in our work, we considered massive MIMO with collocated antenna arrays at the BS. Alternatively, massive BS antenna arrays can be distributed in a large area. Design and analysis of distributed Massive MIMO systems are of interest. Some related works are considered in the literature. For example, in [49], the authors proposed a distributed massive MIMO structure by clustering the cooperating BS and partitioning the users into groups. It was shown that the proposed scheme can achieve a spectral efficiency comparable with that of collocated Massive MIMO in [13] with

a much smaller number of active antennas. However, in [13], the authors considered conjugate beamforming. In future work, the comparison between [49] and [13] with ZF should be considered. It is also interesting to compare the energy efficiency between the system in [49] and the system in [13].

References

[1] Qualcom Inc., *The* 100× *Data Challenge*, Oct. 2013. [Online]. Available: http://www.qualcomm.com/solutions/wireless-networks/technologies/1000x-data.

[2] Ericson, *5G Radio Access—Research and Vision*, June 2013. [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf.

[3] Cisco, "*Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018*," Feb. 2014. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-520862.html.

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

[5] 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.

[6] V. Stankovic and M. Haardt, "Generalized design of multiuser MIMO precoding matrices," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 953–961, Mar. 2008.

[7] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[8] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality" *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

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

[10] S. Verdú, *Multiuser Dectection*. Cambridge, UK: Cambridge University Press, 1998.

[11] N. Jindal and A. Goldsmith, "Dirty-paper coding vs. TDMA for MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

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

[13] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[14] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–46, Jan. 2013.

[15] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[16] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[17] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

[18] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, Aug. 2012.

[19] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 276–279, Aug. 2012.

[20] ——, "Effect of oscillator phase noise on uplink performance of large MU-MIMO systems," in *Proc. of the 50-th Annual Allerton Conference on Communication, Control, and Computing*, 2012.

[21] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channel," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[22] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[23] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 303–313, Feb. 2013.

[24] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Interference reduction on cellular networks with large antenna arrays," in *Proc. IEEE International Conference on Communications (ICC)*, Ottawa, Canada, Jun. 2012.

[25] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, San Francisco, CA, US, Sept. 2011.

[26] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 GHz," in *Proc. 6th European Conference on Antennas and Propagation (EuCAP)*, Prague, Czech Republic, Mar. 2012.

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

[28] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.

[29] K. Nishimori, K. Cho, Y. Takatori, and T. Hori, "Automatic calibration method using transmitting signals of an adaptive array for TDD sys- tems," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1636–1640, 2001.

[30] R. Rogalin, O. Y. Bursalioglu, H. C. Papadopoulos, G. Caire, and A. F. Molisch, "Hardware-impairment compensation for enabling distributed large-scale MIMO," in *Proc. ITA Workshop*, San Diego, CA, USA, 2013.

[31] H. Q. Ngo and E. G. Larsson, "Blind estimation of effective downlink channel gains in Massive MIMO," in *Proc. IEEE International Conference on Acoustics, Speed and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015, submitted.

[32] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[33] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimations for multicell multiuser MIMO with very large antenna arrays," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 2012.

[34] R. Müeller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Sig. Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[35] A. Ashikhmin and T. L. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE International Symposium on Information Theory (ISIT*), Cambridge, MA, Jul. 2012.

[36] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[37] H. Cramér, *Random Variables and Probability Distributions.* Cambridge, UK: Cambridge University Press, 1970.

[38] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Proc. 49th Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, Illinois, US, Sep. 2011.

[39] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'11)*, Prague, Czech Republic, May 2011, pp. 3464–3467.

[40] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Aspects of favorable propagation in Massive MIMO," in *Proc. European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, Sep. 2014.

[41] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Proc. Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, Illinois, US, Oct. 2013.

[42] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Commun. Lett.*, vol. 3, Dec. 2014.

[43] H. Q. Ngo and E. Larsson, "Large-Scale Multipair Two-Way Relay Networks with Distributed AF Beamforming," *IEEE Commun. Lett.*,vol. 17, no. 12, pp. 2288–2291, Dec. 2013.

[44] H. Q. Ngo and E. G. Larsson, "Spectral efficiency of the multi-pair two-way relay channel with massive arrays," in *Proc. Asilomar Conference on Signals, Systems, and Computer*, Parific Grove, CA, Nov. 2013.

[45] Hien Quoc Ngo, Himal A. Suraweera, Michail Matthaiou, and Erik G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing", *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.

[46] H. Q. Ngo, E. Larsson, and T. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[47] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[48] 3GPP TR 36.211 V 11.1.0 (Release 11), "Evovled universal terrestrial radio access (E-UTRA)," Dec. 2012.

[49] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving "Massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.

Part II

Fundamentals of Massive MIMO

Paper A

Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems

Refereed article published in the IEEE Transactions on
Communications 2013.

Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems

Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta

Abstract

A multiplicity of autonomous terminals simultaneously transmits data streams to a compact array of antennas. The array uses imperfect channel-state information derived from transmitted pilots to extract the individual data streams. The power radiated by the terminals can be made inversely proportional to the square-root of the number of base station antennas with no reduction in performance. In contrast if perfect channel-state information were available the power could be made inversely proportional to the number of antennas. Lower capacity bounds for maximum-ratio combining (MRC), zero-forcing (ZF) and minimum mean-square error (MMSE) detection are derived. An MRC receiver normally performs worse than ZF and MMSE. However as power levels are reduced, the cross-talk introduced by the inferior maximum-ratio receiver eventually falls below the noise level and this simple receiver becomes a viable option. The tradeoff between the energy efficiency (as measured in bits/J) and spectral efficiency (as measured in bits/channel use/terminal) is quantified for a channel model that includes small-scale fading but not large-scale fading. It is shown that the use of moderately large antenna arrays can improve the spectral and energy efficiency with orders of magnitude compared to a single-antenna system.

1 Introduction

In multiuser multiple-input multiple-output (MU-MIMO) systems, a base station (BS) equipped with multiple antennas serves a number of users. Such systems have attracted much attention for some time now [1]. Conventionally, the communication between the BS and the users is performed by orthogonalizing the channel so that the BS communicates with each user in separate time-frequency resources. This is not optimal from an information-theoretic point of view, and higher rates can be achieved if the BS communicates with several users in the same time-frequency resource [2,3]. However, complex techniques to mitigate inter-user interference must then be used, such as maximum-likelihood multiuser detection on the uplink [4], or "dirty-paper coding" on the downlink [5,6].

Recently, there has been a great deal of interest in MU-MIMO with *very large antenna arrays* at the BS. Very large arrays can substantially reduce intracell interference with simple signal processing [7]. We refer to such systems as "very large MU-MIMO systems" here, and with very large we mean arrays comprising say a hundred, or a few hundreds, of antennas, simultaneously serving tens of users. The design and analysis of very large MU-MIMO systems is a fairly new subject that is attracting substantial interest [7–10]. The vision is that each individual antenna can have a small physical size, and be built from inexpensive hardware. With a very large antenna array, things that were random before start to look deterministic. As a consequence, the effect of small-scale fading can be averaged out. Furthermore, when the number of BS antennas grows large, the random channel vectors between the users and the BS become pairwisely orthogonal [9]. In the limit of an infinite number of antennas, with simple matched filter processing at the BS, uncorrelated noise and intracell interference disappear completely [7]. Another important advantage of large MIMO systems is that they enable us to reduce the transmitted power. On the uplink, reducing the transmit power of the terminals will drain their batteries slower. On the downlink, much of the electrical power consumed by a BS is spent by power amplifiers and associated circuits and cooling systems [11]. Hence reducing the emitted RF power would help in cutting the electricity consumption of the BS.

This paper analyzes the potential for power savings on the uplink of very large MU-MIMO systems. We derive new capacity bounds of the uplink for finite number of BS antennas. While it is well known that MIMO technology can offer improved power efficiency, owing to both array gains and diversity effects [12], we are not aware of any work that analyzes power efficiency of MU-MIMO systems with receiver structures that are realistic for very large MIMO.We consider both single-cell and multicell systems, but focus on the analysis of single-cell MU-MIMO systems since: i) the results are easily comprehensible; ii) it bounds the performance of a multicell system; and iii) the single-cell performance can be actually attained if one uses successively less-aggressive frequency-reuse (e.g., with reuse factor 3, or 7). Our results are different from recent results in [13] and [14]. In [13] and [14], the authors

derived a deterministic equivalent of the SINR assuming that the number of transmit antennas and the number of users go to infinity but their ratio remains bounded for the downlink of network MIMO systems using a sophisticated scheduling scheme and MISO broadcast channels using zero-forcing (ZF) precoding, respectively.

The paper makes the following specific contributions:

- We show that, when the number of BS antennas M grows without bound, we can reduce the transmitted power of each user proportionally to $1/M$ if the BS has perfect channel state information (CSI), and proportionally to $1/\sqrt{M}$ if CSI is estimated from uplink pilots. This holds true even when using simple, linear receivers. We also derive closed-form expressions of lower bounds on the uplink achievable rates for finite M, for the cases of perfect and imperfect CSI, assuming MRC, ZF, and minimum mean-squared error (MMSE) receivers, respectively. See Section 3.
- We study the tradeoff between spectral efficiency and energy efficiency. For imperfect CSI, in the low transmit power regime, we can simultaneously increase the spectral-efficiency and energy-efficiency. We further show that in large-scale MIMO, very high spectral efficiency can be obtained even with simple MRC processing at the same time as the transmit power can be cut back by orders of magnitude and that this holds true even when taking into account the losses associated with acquiring CSI from uplink pilots. MRC also has the advantage that it can be implemented in a distributed manner, i.e., each antenna performs multiplication of the received signals with the conjugate of the channel, without sending the entire baseband signal to the BS for processing. Quantitatively, our energy-spectral efficiency tradeoff analysis incorporates the effects of small-scale fading but neglects those of large-scale fading, leaving an analysis of the effect of large-scale fading for future work. See Section 4.

2 System Model and Preliminaries

2.1 MU-MIMO System Model

We consider the uplink of a MU-MIMO system. The system includes one BS equipped with an array of M antennas that receive data from K single-antenna users. The nice thing about single-antenna users is that they are inexpensive, simple, and power-efficient, and each user still gets typically high throughput. Furthermore, the assumption that users have single antennas can be considered as a special case of users having multiple antennas when we treat the extra antennas as

if they were additional autonomous users.[1] The users transmit their data in the same time-frequency resource. The $M \times 1$ received vector at the BS is

$$\boldsymbol{y} = \sqrt{p_{\mathrm{u}}}\boldsymbol{G}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{G}$ represents the $M \times K$ channel matrix between the BS and the K users, i.e., $g_{mk} \triangleq [\boldsymbol{G}]_{mk}$ is the channel coefficient between the mth antenna of the BS and the kth user; $\sqrt{p_{\mathrm{u}}}\boldsymbol{x}$ is the $K \times 1$ vector of symbols simultaneously transmitted by the K users (the average transmitted power of each user is p_{u}); and $\boldsymbol{n}$ is a vector of additive white, zero-mean Gaussian noise. We take the noise variance to be 1, to minimize notation, but without loss of generality. With this convention, p_{u} has the interpretation of normalized "transmit" SNR and is therefore dimensionless. The model (3) also applies to wideband channels handled by OFDM over restricted intervals of frequency.

The channel matrix $\boldsymbol{G}$ models independent fast fading, geometric attenuation, and log-normal shadow fading. The coefficient g_{mk} can be written as

$$g_{mk} = h_{mk}\sqrt{\beta_k}, \quad m = 1, 2, ..., M, \tag{2}$$

where h_{mk} is the fast fading coefficient from the kth user to the mth antenna of the BS. $\sqrt{\beta_k}$ models the geometric attenuation and shadow fading which is assumed to be independent over m and to be constant over many coherence time intervals and known a priori. This assumption is reasonable since the distances between the users and the BS are much larger than the distance between the antennas, and the value of β_k changes very slowly with time. Then, we have

$$\boldsymbol{G} = \boldsymbol{H}\boldsymbol{D}^{1/2}, \tag{3}$$

where $\boldsymbol{H}$ is the $M \times K$ matrix of fast fading coefficients between the K users and the BS, i.e., $[\boldsymbol{H}]_{mk} = h_{mk}$, and $\boldsymbol{D}$ is a $K \times K$ diagonal matrix, where $[\boldsymbol{D}]_{kk} = \beta_k$. Therefore, (3) can be written as

$$\boldsymbol{y} = \sqrt{p_{\mathrm{u}}}\boldsymbol{H}\boldsymbol{D}^{1/2}\boldsymbol{x} + \boldsymbol{n}. \tag{4}$$

2.2 Review of Some Results on Very Long Random Vectors

We review some limit results for random vectors [15] that will be useful later on. Let $\boldsymbol{p} \triangleq [p_1 \ ... \ p_n]^T$ and $\boldsymbol{q} \triangleq [q_1 \ ... \ q_n]^T$ be mutually independent $n \times 1$ vectors

[1] Note that under the assumptions on favorable propagation (see Section 2.3), having n autonomous single-antenna users or having one n-antenna user (where the antennas cooperate in the encoding), represent two cases with equal energy and spectral efficiency. To see why, consider two cases: the case of 2 autonomous single-antenna users of which each spends power P, and the case of one dual-antenna user with a total power constraint of $2P$. Then, the sum rates for the two cases are the same and equal to $\log_2\left(1 + \frac{P\|\boldsymbol{h}_1\|^2}{N_0}\right) + \log_2\left(1 + \frac{P\|\boldsymbol{h}_2\|^2}{N_0}\right) = \log_2 \det\left(\boldsymbol{I} + \frac{1}{N_0}[\boldsymbol{h}_1 \ \boldsymbol{h}_2]\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \boldsymbol{h}_1^H \\ \boldsymbol{h}_2^H \end{bmatrix}\right)$, where $\boldsymbol{h}_i$ is the channel vector between the ith user (or ith antenna) to the base station, and N_0 is the variance of noise.

whose elements are i.i.d. zero-mean random variables (RVs) with $\mathbb{E}\left\{|p_i|^2\right\} = \sigma_p^2$, and $\mathbb{E}\left\{|q_i|^2\right\} = \sigma_q^2$, $i = 1, ..., n$. Then from the law of large numbers,

$$\frac{1}{n}\boldsymbol{p}^H\boldsymbol{p} \xrightarrow{a.s.} \sigma_p^2, \text{ and } \frac{1}{n}\boldsymbol{p}^H\boldsymbol{q} \xrightarrow{a.s.} 0, \text{ as } n \to \infty. \tag{5}$$

where $\xrightarrow{a.s.}$ denotes the almost sure convergence. Also, from the Lindeberg-Lévy central limit theorem,

$$\frac{1}{\sqrt{n}}\boldsymbol{p}^H\boldsymbol{q} \xrightarrow{d} \mathcal{CN}\left(0, \sigma_p^2\sigma_q^2\right), \text{ as } n \to \infty, \tag{6}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

2.3 Favorable Propagation

Throughout the rest of the paper, we assume that the fast fading coefficients, i.e., the elements of $\boldsymbol{H}$ are i.i.d. RVs with zero mean and unit variance. Then the conditions in (5)–(6) are satisfied with $\boldsymbol{p}$ and $\boldsymbol{q}$ being any two distinct columns of $\boldsymbol{G}$. In this case we have

$$\frac{\boldsymbol{G}^H\boldsymbol{G}}{M} = \boldsymbol{D}^{1/2}\frac{\boldsymbol{H}^H\boldsymbol{H}}{M}\boldsymbol{D}^{1/2} \approx \boldsymbol{D}, \quad M \gg K,$$

and we say that we have *favorable propagation*. Clearly, if all fading coefficients are i.i.d. and zero mean, we have favorable propagation. Recent channel measurements campaigns have shown that multiuser MIMO systems with large antenna arrays have characteristics that approximate the favorable-propagation assumption fairly well [9], and therefore provide experimental justification for this assumption.

To understand why favorable propagation is desirable, consider an $M \times K$ uplink (multiple-access) MIMO channel $\boldsymbol{H}$, where $M \geq K$, neglecting for now path loss and shadowing factors in $\boldsymbol{D}$. This channel can offer a sum-rate of

$$R = \sum_{k=1}^{K} \log_2\left(1 + p_{\mathrm{u}}\lambda_k^2\right), \tag{7}$$

where p_{u} is the power spent per terminal and $\{\lambda_k\}_{k=1}^K$ are the singular values of $\boldsymbol{H}$, see [12]. If the channel matrix is normalized such that $|H_{ij}| \sim 1$ (where $\sim$ means equality of the order of magnitude), then $\sum_{k=1}^K \lambda_k^2 = \|\boldsymbol{H}\|^2 \approx MK$. Under this constraint the rate R is bounded as

$$\log_2\left(1 + MKp_{\mathrm{u}}\right) \leq R \leq K\log_2\left(1 + Mp_{\mathrm{u}}\right). \tag{8}$$

The lower bound (left inequality) is satisfied with equality if $\lambda_1^2 = MK$ and $\lambda_2^2 = \cdots = \lambda_K^2 = 0$ and corresponds to a rank-one (line-of-sight) channel. The upper bound (right inequality) is achieved if $\lambda_1^2 = \cdots = \lambda_K^2 = M$. This occurs if the columns of $\boldsymbol{H}$ are mutually orthogonal and have the same norm, which is the case when we have favorable propagation.

3 Achievable Rate and Asymptotic ($M \to \infty$) Power Efficiency

By using a large antenna array, we can reduce the transmitted power of the users as M grows large, while maintaining a given, desired quality-of-service. In this section, we quantify this potential for power decrease, and derive achievable rates of the uplink. Theoretically, the BS can use the maximum-likelihood detector to obtain optimal performance. However, the complexity of this detector grows exponentially with K. The interesting operating regime is when both M and K are large, but M is still (much) larger than K, i.e., $1 \ll K \ll M$. It is known that in this case, linear detectors (MRC, ZF and MMSE) perform fairly well [7] and therefore we will restrict consideration to those detectors in this paper. We treat the cases of perfect CSI (Section 3.1) and estimated CSI (Section 3.2) separately.

3.1 Perfect Channel State Information

We first consider the case when the BS has perfect CSI, i.e. it knows $\boldsymbol{G}$. Let $\boldsymbol{A}$ be an $M \times K$ linear detector matrix which depends on the channel $\boldsymbol{G}$. By using the linear detector, the received signal is separated into streams by multiplying it with $\boldsymbol{A}^H$ as follows

$$\boldsymbol{r} = \boldsymbol{A}^H \boldsymbol{y}. \tag{9}$$

We consider three conventional linear detectors MRC, ZF, and MMSE, i.e.,

$$\boldsymbol{A} = \begin{cases} \boldsymbol{G} & \text{for MRC} \\ \boldsymbol{G}\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1} & \text{for ZF} \\ \boldsymbol{G}\left(\boldsymbol{G}^H\boldsymbol{G} + \frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_K\right)^{-1} & \text{for MMSE.} \end{cases} \tag{10}$$

From (3) and (9), the received vector after using the linear detector is given by

$$\boldsymbol{r} = \sqrt{p_{\mathrm{u}}}\boldsymbol{A}^H\boldsymbol{G}\boldsymbol{x} + \boldsymbol{A}^H\boldsymbol{n}. \tag{11}$$

Let r_k and x_k be the kth elements of the $K \times 1$ vectors $\boldsymbol{r}$ and $\boldsymbol{x}$, respectively. Then,

$$r_k = \sqrt{p_{\mathrm{u}}}\boldsymbol{a}_k^H\boldsymbol{G}\boldsymbol{x} + \boldsymbol{a}_k^H\boldsymbol{n} = \sqrt{p_{\mathrm{u}}}\boldsymbol{a}_k^H\boldsymbol{g}_k x_k + \sqrt{p_{\mathrm{u}}}\sum_{i=1, i\neq k}^{K} \boldsymbol{a}_k^H\boldsymbol{g}_i x_i + \boldsymbol{a}_k^H\boldsymbol{n}, \tag{12}$$

where $\boldsymbol{a}_k$ and $\boldsymbol{g}_k$ are the kth columns of the matrices $\boldsymbol{A}$ and $\boldsymbol{G}$, respectively. For a fixed channel realization $\boldsymbol{G}$, the noise-plus-interference term is a random variable with zero mean and variance $p_{\mathrm{u}}\sum_{i=1, i\neq k}^{K} |\boldsymbol{a}_k^H\boldsymbol{g}_i|^2 + \|\boldsymbol{a}_k\|^2$. By modeling this term

as additive Gaussian noise independent of x_k we can obtain a lower bound on the achievable rate. Assuming further that the channel is ergodic so that each codeword spans over a large (infinite) number of realizations of the fast-fading factor of $\boldsymbol{G}$, the ergodic achievable uplink rate of the kth user is

$$R_{\mathrm{P},k} = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{u}}|\boldsymbol{a}_k^H \boldsymbol{g}_k|^2}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\boldsymbol{a}_k^H \boldsymbol{g}_i|^2 + \|\boldsymbol{a}_k\|^2}\right)\right\}. \tag{13}$$

To approach this capacity lower bound, the message has to be encoded over many realizations of all sources of randomness that enter the model (noise and channel). In practice, assuming wideband operation, this can be achieved by coding over the frequency domain, using, for example coded OFDM.

Proposition 1 *Assume that the BS has perfect CSI and that the transmit power of each user is scaled with M according to $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{M}$, where E_{u} is fixed. Then,*[2]

$$R_{\mathrm{P},k} \to \log_2\left(1 + \beta_k E_{\mathrm{u}}\right), M \to \infty. \tag{14}$$

Proof: *We give the proof for the case of an MRC receiver. With MRC, $\boldsymbol{A} = \boldsymbol{G}$ so $\boldsymbol{a}_k = \boldsymbol{g}_k$. From (13), the achievable uplink rate of the kth user is*

$$R_{\mathrm{P},k}^{\mathtt{mrc}} = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^4}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\boldsymbol{g}_k^H \boldsymbol{g}_i|^2 + \|\boldsymbol{g}_k\|^2}\right)\right\}. \tag{15}$$

Substituting $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{M}$ into (15), and using (5), we obtain (14). By using the law of large numbers, we can arrive at the same result for the ZF and MMSE receivers. Note from (3) and (5) that when M grows large, $\frac{1}{M}\boldsymbol{G}^H\boldsymbol{G}$ tends to $\boldsymbol{D}$, and hence the ZF and MMSE filters tend to that of the MRC. □

Proposition 1 shows that with perfect CSI at the BS and a large M, the performance of a MU-MIMO system with M antennas at the BS and a transmit power per user of E_{u}/M is equal to the performance of a SISO system with transmit power E_{u}, without any intra-cell interference and without any fast fading. In other words, by using a large number of BS antennas, we can scale down the transmit power proportionally to $1/M$. At the same time we increase the spectral efficiency K times by simultaneously serving K users in the same time-frequency resource.

[2] As mentioned after (3), p_{u} has the interpretation of normalized transmit SNR, and it is dimensionless. Therefore E_{u} is dimensionless too.

3.1.1 Maximum-Ratio Combining

For MRC, from (15), by the convexity of $\log_2\left(1+\frac{1}{x}\right)$ and using Jensen's inequality, we obtain the following lower bound on the achievable rate:

$$R_{\mathrm{P},k}^{\mathtt{mrc}} \geq \tilde{R}_{\mathrm{P},k}^{\mathtt{mrc}} \triangleq \log_2\left(1+\left(\mathbb{E}\left\{\frac{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\boldsymbol{g}_k^H\boldsymbol{g}_i|^2+\|\boldsymbol{g}_k\|^2}{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^4}\right\}\right)^{-1}\right). \tag{16}$$

Proposition 2 *With perfect CSI, Rayleigh fading, and $M \geq 2$, the uplink achievable rate from the kth user for MRC can be lower bounded as follows:*

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mrc}} = \log_2\left(1+\frac{p_{\mathrm{u}}\left(M-1\right)\beta_k}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}\beta_i+1}\right). \tag{17}$$

Proof: *See Appendix A.* □

If $p_{\mathrm{u}} = E_{\mathrm{u}}/M$, and M grows without bound, then from (17), we have

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mrc}} = \log_2\left(1+\frac{\frac{E_{\mathrm{u}}}{M}\left(M-1\right)\beta_k}{\frac{E_{\mathrm{u}}}{M}\sum_{i=1,i\neq k}^{K}\beta_i+1}\right) \rightarrow \log_2\left(1+\beta_k E_{\mathrm{u}}\right),\ M\rightarrow\infty. \tag{18}$$

Equation (18) shows that the lower bound in (17) becomes equal to the exact limit in Proposition 1 as $M \rightarrow \infty$.

3.1.2 Zero-Forcing Receiver

With ZF, $\boldsymbol{A}^H = \left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H$, or $\boldsymbol{A}^H\boldsymbol{G} = \boldsymbol{I}_K$. Therefore, $\boldsymbol{a}_k^H\boldsymbol{g}_i = \delta_{ki}$, where $\delta_{ki} = 1$ when $k = i$ and 0 otherwise. From (13), the uplink rate for the kth user is

$$R_{\mathrm{P},k}^{\mathtt{zf}} = \mathbb{E}\left\{\log_2\left(1+\frac{p_{\mathrm{u}}}{\left[\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}}\right)\right\}. \tag{19}$$

By using Jensen's inequality, we obtain the following lower bound on the achievable rate:

$$R_{\mathrm{P},k}^{\mathtt{zf}} \geq \tilde{R}_{\mathrm{P},k}^{\mathtt{zf}} = \log_2\left(1+\frac{p_{\mathrm{u}}}{\mathbb{E}\left\{\left[\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}\right\}}\right). \tag{20}$$

Proposition 3 *When using ZF, in Rayleigh fading, and provided that $M \geq K+1$, the achievable uplink rate for the kth user is lower bounded by*

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{zf}} = \log_2\left(1 + p_{\mathrm{u}}\left(M-K\right)\beta_k\right). \tag{21}$$

Proof: *See Appendix B.* □

If $p_{\mathrm{u}} = E_{\mathrm{u}}/M$, and M grows large, we have

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{zf}} = \log_2\left(1 + \frac{E_{\mathrm{u}}}{M}\left(M-K\right)\beta_k\right) \to \log_2\left(1+\beta_k E_{\mathrm{u}}\right), \ M \to \infty. \tag{22}$$

We can see again from (22) that the lower bound becomes exact for large M.

3.1.3 Minimum Mean-Squared Error Receiver

For MMSE, the detector matrix $\boldsymbol{A}$ is

$$\boldsymbol{A}^H = \left(\boldsymbol{G}^H\boldsymbol{G} + \frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_K\right)^{-1}\boldsymbol{G}^H = \boldsymbol{G}^H\left(\boldsymbol{G}\boldsymbol{G}^H + \frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M\right)^{-1}. \tag{23}$$

Therefore, the kth column of $\boldsymbol{A}$ is given by [16]

$$\boldsymbol{a}_k = \left(\boldsymbol{G}\boldsymbol{G}^H + \frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M\right)^{-1}\boldsymbol{g}_k = \frac{\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k}{\boldsymbol{g}_k^H\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k + 1}, \tag{24}$$

where $\boldsymbol{\Lambda}_k \triangleq \sum_{i=1, i\neq k}^{K}\boldsymbol{g}_i\boldsymbol{g}_i^H + \frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M$. Substituting (24) into (13), we obtain the uplink rate for user k:

$$\begin{aligned}
R_{\mathrm{P},k}^{\mathtt{mmse}} &= \mathbb{E}\left\{\log_2\left(1+\boldsymbol{g}_k^H\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k\right)\right\} \\
&\overset{(a)}{=} \mathbb{E}\left\{\log_2\left(\frac{1}{1-\boldsymbol{g}_k^H\left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M + \boldsymbol{G}\boldsymbol{G}^H\right)^{-1}\boldsymbol{g}_k}\right)\right\} \\
&= \mathbb{E}\left\{\log_2\left(\frac{1}{1-\left[\boldsymbol{G}^H\left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M + \boldsymbol{G}\boldsymbol{G}^H\right)^{-1}\boldsymbol{G}\right]_{kk}}\right)\right\} \\
&\overset{(b)}{=} \mathbb{E}\left\{\log_2\left(\frac{1}{\left[\left(\boldsymbol{I}_K + p_{\mathrm{u}}\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}}\right)\right\},
\end{aligned} \tag{25}$$

where (a) is obtained directly from (24), and (b) is obtained by using the identity

$$\boldsymbol{G}^H\left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M+\boldsymbol{G}\boldsymbol{G}^H\right)^{-1}\boldsymbol{G}=\left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_K+\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\boldsymbol{G}=\boldsymbol{I}_K-\left(\boldsymbol{I}_K+p_{\mathrm{u}}\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}.$$

By using Jensen's inequality, we obtain the following lower bound on the achievable uplink rate:

$$R_{\mathrm{P},k}^{\mathtt{mmse}} \geq \tilde{R}_{\mathrm{P},k}^{\mathtt{mmse}} = \log_2\left(1+\frac{1}{\mathbb{E}\left\{1/\gamma_k\right\}}\right), \tag{26}$$

where $\gamma_k = \frac{1}{\left[\left(\boldsymbol{I}_K+p_{\mathrm{u}}\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}} - 1$. For Rayleigh fading, the exact distribution of γ_k can be found in [17]. This distribution is analytically intractable. To proceed, we approximate it with a distribution which has an analytically tractable form. More specifically, the PDF of γ_k can be approximated by a Gamma distribution as follows [18]:

$$p_{\gamma_k}(\gamma) = \frac{\gamma^{\alpha_k-1}e^{-\gamma/\theta_k}}{\Gamma(\alpha_k)\,\theta_k^{\alpha_k}}, \tag{27}$$

where

$$\alpha_k = \frac{\left(M-K+1+(K-1)\,\mu\right)^2}{M-K+1+(K-1)\,\kappa},\ \theta_k = \frac{M-K+1+(K-1)\,\kappa}{M-K+1+(K-1)\,\mu}p_{\mathrm{u}}\beta_k, \tag{28}$$

and where μ and κ are determined by solving following equations:

$$\begin{aligned}\mu &= \frac{1}{K-1}\sum_{i=1,i\neq k}^{K}\frac{1}{Mp_{\mathrm{u}}\beta_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\mu\right)+1}\\ \kappa&\left(1+\sum_{i=1,i\neq k}^{K}\frac{p_{\mathrm{u}}\beta_i}{\left(Mp_{\mathrm{u}}\beta_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\mu\right)+1\right)^2}\right)\\ &= \sum_{i=1,i\neq k}^{K}\frac{p_{\mathrm{u}}\beta_i\mu+1}{\left(Mp_{\mathrm{u}}\beta_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\mu\right)+1\right)^2}.\end{aligned} \tag{29}$$

Using the approximate PDF of γ_k given by (27), we have the following proposition.

Proposition 4 *With perfect CSI, Rayleigh fading, and MMSE, the lower bound on the achievable rate for the kth user can be approximated as*

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mmse}} = \log_2\left(1+(\alpha_k-1)\,\theta_k\right). \tag{30}$$

Proof: *Substituting* (27) *into* (26)*, and using the identity [19, eq. (3.326.2)], we obtain*

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mmse}} = \log_2\left(1 + \frac{\Gamma(\alpha_k)}{\Gamma(\alpha_k - 1)}\theta_k\right), \tag{31}$$

where $\Gamma(\cdot)$ *is the Gamma function. Then, using* $\Gamma(x+1) = x\Gamma(x)$*, we obtain the desired result* (30). □

Remark 1 *From* (13)*, the achievable rate* $R_{\mathrm{P},k}$ *can be rewritten as*

$$\begin{aligned} R_{\mathrm{P},k} = \mathbb{E}\left\{\log_2\left(1 + \frac{|\boldsymbol{a}_k^H \boldsymbol{g}_k|^2}{\boldsymbol{a}_k^H \boldsymbol{\Lambda}_k \boldsymbol{a}_k}\right)\right\} &\leq \mathbb{E}\left\{\log_2\left(1 + \frac{\|\boldsymbol{a}_k^H \boldsymbol{\Lambda}_k^{1/2}\|^2 \|\boldsymbol{\Lambda}_k^{-1/2}\boldsymbol{g}_k\|^2}{\boldsymbol{a}_k^H \boldsymbol{\Lambda}_k \boldsymbol{a}_k}\right)\right\} \\ &= \mathbb{E}\left\{\log_2\left(1 + \boldsymbol{g}_k^H \boldsymbol{\Lambda}_k^{-1} \boldsymbol{g}_k\right)\right\}. \end{aligned} \tag{32}$$

The inequality is obtained by using Cauchy-Schwarz' inequality, which holds with equality when $\boldsymbol{a}_k = c\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k$*, for any* $c \in \mathbb{C}$*. This corresponds to the MMSE detector (see* (24)*). This implies that the MMSE detector is optimal in the sense that it maximizes the achievable rate given by* (13)*.*

3.2 Imperfect Channel State Information

In practice, the channel matrix $\boldsymbol{G}$ has to be estimated at the BS. The standard way of doing this is to use uplink pilots. A part of the coherence interval of the channel is then used for the uplink training. Let T be the length (time-bandwidth product) of the coherence interval and let τ be the number of symbols used for pilots. During the training part of the coherence interval, all users simultaneously transmit mutually orthogonal pilot sequences of length τ symbols. The pilot sequences used by the K users can be represented by a $\tau \times K$ matrix $\sqrt{p_{\mathrm{p}}}\boldsymbol{\Phi}$ ($\tau \geq K$), which satisfies $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{I}_K$, where $p_{\mathrm{p}} \triangleq \tau p_{\mathrm{u}}$. Then, the $M \times \tau$ received pilot matrix at the BS is given by

$$\boldsymbol{Y}_{\mathrm{p}} = \sqrt{p_{\mathrm{p}}}\boldsymbol{G}\boldsymbol{\Phi}^T + \boldsymbol{N}, \tag{33}$$

where $\boldsymbol{N}$ is an $M \times \tau$ matrix with i.i.d. $\mathcal{CN}(0,1)$ elements. The MMSE estimate of $\boldsymbol{G}$ given $\boldsymbol{Y}$ is

$$\hat{\boldsymbol{G}} = \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{Y}_{\mathrm{p}}\boldsymbol{\Phi}^*\tilde{\boldsymbol{D}} = \left(\boldsymbol{G} + \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{W}\right)\tilde{\boldsymbol{D}}, \tag{34}$$

where $\boldsymbol{W} \triangleq \boldsymbol{N}\boldsymbol{\Phi}^*$, and $\tilde{\boldsymbol{D}} \triangleq \left(\frac{1}{p_{\mathrm{p}}}\boldsymbol{D}^{-1} + \boldsymbol{I}_K\right)^{-1}$. Since $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{I}_K$, $\boldsymbol{W}$ has i.i.d. $\mathcal{CN}(0,1)$ elements. Note that our analysis takes into account the fact that pilot signals cannot take advantage of the large number of receive antennas since channel

estimation has to be done on a per-receive antenna basis. All results that we present take this fact into account. Denote by $\boldsymbol{\mathcal{E}} \triangleq \hat{\boldsymbol{G}} - \boldsymbol{G}$. Then, from (34), the elements of the ith column of $\boldsymbol{\mathcal{E}}$ are RVs with zero means and variances $\frac{\beta_i}{p_{\mathrm{p}}\beta_i+1}$. Furthermore, owing to the properties of MMSE estimation, $\boldsymbol{\mathcal{E}}$ is independent of $\hat{\boldsymbol{G}}$. The received vector at the BS can be rewritten as

$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{A}}^H \left(\sqrt{p_{\mathrm{u}}} \hat{\boldsymbol{G}} \boldsymbol{x} - \sqrt{p_{\mathrm{u}}} \boldsymbol{\mathcal{E}} \boldsymbol{x} + \boldsymbol{n} \right). \tag{35}$$

Therefore, after using the linear detector, the received signal associated with the kth user is

$$\begin{aligned} \hat{r}_k &= \sqrt{p_{\mathrm{u}}} \hat{\boldsymbol{a}}_k^H \hat{\boldsymbol{G}} \boldsymbol{x} - \sqrt{p_{\mathrm{u}}} \hat{\boldsymbol{a}}_k^H \boldsymbol{\mathcal{E}} \boldsymbol{x} + \hat{\boldsymbol{a}}_k^H \boldsymbol{n} \\ &= \sqrt{p_{\mathrm{u}}} \hat{\boldsymbol{a}}_k^H \hat{\boldsymbol{g}}_k x_k + \sqrt{p_{\mathrm{u}}} \sum_{i=1, i \neq k}^{K} \hat{\boldsymbol{a}}_k^H \hat{\boldsymbol{g}}_i x_i - \sqrt{p_{\mathrm{u}}} \sum_{i=1}^{K} \hat{\boldsymbol{a}}_k^H \boldsymbol{\varepsilon}_i x_i + \hat{\boldsymbol{a}}_k^H \boldsymbol{n}, \end{aligned} \tag{36}$$

where $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{g}}_i$, and $\boldsymbol{\varepsilon}_i$ are the ith columns of $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{G}}$, and $\boldsymbol{\mathcal{E}}$, respectively.

Since $\hat{\boldsymbol{G}}$ and $\boldsymbol{\mathcal{E}}$ are independent, $\hat{\boldsymbol{A}}$ and $\boldsymbol{\mathcal{E}}$ are independent too. The BS treats the channel estimate as the true channel, and the part including the last three terms of (36) is considered as interference and noise. Therefore, an achievable rate of the uplink transmission from the kth user is given by

$$R_{\mathrm{IP},k} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{p_{\mathrm{u}} |\hat{\boldsymbol{a}}_k^H \hat{\boldsymbol{g}}_k|^2}{p_{\mathrm{u}} \sum_{i=1, i \neq k}^{K} |\hat{\boldsymbol{a}}_k^H \hat{\boldsymbol{g}}_i|^2 + p_{\mathrm{u}} \|\hat{\boldsymbol{a}}_k\|^2 \sum_{i=1}^{K} \frac{\beta_i}{\tau p_{\mathrm{u}} \beta_i + 1} + \|\hat{\boldsymbol{a}}_k\|^2} \right) \right\}. \tag{37}$$

Intuitively, if we cut the transmitted power of each user, both the data signal and the pilot signal suffer from the reduction in power. Since these signals are multiplied together at the receiver, we expect that there will be a "squaring effect". As a consequence, we cannot reduce power proportionally to $1/M$ as in the case of perfect CSI. The following proposition shows that it is possible to reduce the power (only) proportionally to $1/\sqrt{M}$.

Proposition 5 *Assume that the BS has imperfect CSI, obtained by MMSE estimation from uplink pilots, and that the transmit power of each user is* $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{\sqrt{M}}$, *where* E_{u} *is fixed. Then,*

$$R_{\mathrm{IP},k} \to \log_2 \left(1 + \tau \beta_k^2 E_{\mathrm{u}}^2 \right), M \to \infty. \tag{38}$$

Proof: *For MRC, substituting* $\hat{\boldsymbol{a}}_k = \hat{\boldsymbol{g}}_k$ *into* (37), *we obtain the achievable uplink rate as*

$$R_{\mathrm{IP},k}^{\mathtt{mrc}} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{p_{\mathrm{u}} \|\hat{\boldsymbol{g}}_k\|^4}{p_{\mathrm{u}} \sum_{i=1, i \neq k}^{K} |\hat{\boldsymbol{g}}_k^H \hat{\boldsymbol{g}}_i|^2 + p_{\mathrm{u}} \|\hat{\boldsymbol{g}}_k\|^2 \sum_{i=1}^{K} \frac{\beta_i}{\tau p_{\mathrm{u}} \beta_i + 1} + \|\hat{\boldsymbol{g}}_k\|^2} \right) \right\}. \tag{39}$$

Substituting $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$ *into* (39)*, and again using* (5) *along with the fact that each element of* $\hat{\boldsymbol{g}}_k$ *is a RV with zero mean and variance* $\frac{p_{\mathrm{p}}\beta_k^2}{p_{\mathrm{p}}\beta_k+1}$*, we obtain* (38)*. We can obtain the limit in* (38) *for ZF and MMSE in a similar way.* □

Proposition 5 implies that with imperfect CSI and a large M, the performance of a MU-MIMO system with an M-antenna array at the BS and with the transmit power per user set to $E_{\mathrm{u}}/\sqrt{M}$ is equal to the performance of an interference-free SISO link with transmit power $\tau\beta_k E_{\mathrm{u}}^2$, without fast fading.

Remark 2 *From the proof of Proposition 5, we see that if we cut the transmit power proportionally to* $1/M^{\alpha}$*, where* $\alpha > 1/2$*, then the SINR of the uplink transmission from the kth user will go to zero as* $M \to \infty$*. This means that* $1/\sqrt{M}$ *is the fastest rate at which we can cut the transmit power of each user and still maintain a fixed rate.*

Remark 3 *In general, each user can use different transmit powers which depend on the geometric attenuation and the shadow fading. This can be done by assuming that the kth user knows* β_k *and performs power control. In this case, the reasoning leading to Proposition 5 can be extended to show that to achieve the same rate as in a SISO system using transmit power* E_{u}*, we must choose the transmit power of the kth user to be* $\sqrt{\frac{E_{\mathrm{u}}}{M\tau\beta_k}}$*.*

Remark 4 *It can be seen directly from* (15) *and* (39) *that the power-scaling laws still hold even for the most unfavorable propagation case (where* $\boldsymbol{H}$ *has rank one). However, for this case, the multiplexing gains do not materialize since the intracell interference cannot be cancelled when* M *grows without bound.*

3.2.1 Maximum-Ratio Combining

By following a similar line of reasoning as in the case of perfect CSI, we can obtain lower bounds on the achievable rate.

Proposition 6 *With imperfect CSI, Rayleigh fading, MRC processing, and for* $M \geq 2$*, the achievable uplink rate for the kth user is lower bounded by*

$$\tilde{R}_{\mathrm{IP},k}^{\mathtt{mrc}} = \log_2\left(1 + \frac{\tau p_{\mathrm{u}}^2 (M-1)\beta_k^2}{p_{\mathrm{u}}(\tau p_{\mathrm{u}}\beta_k + 1)\sum_{i=1,i\neq k}^{K}\beta_i + (\tau+1)p_{\mathrm{u}}\beta_k + 1}\right). \tag{40}$$

By choosing $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$, we obtain

$$\tilde{R}_{\mathrm{IP},k}^{\mathtt{mrc}} \to \log_2\left(1 + \tau\beta_k^2 E_{\mathrm{u}}^2\right),\ M \to \infty. \tag{41}$$

Again, when $M \to \infty$, the asymptotic bound on the rate equals the exact limit obtained from Proposition 5.

3.2.2 ZF Receiver

For the ZF receiver, we have $\hat{\boldsymbol{a}}_k^H \hat{\boldsymbol{g}}_i = \delta_{ki}$. From (37), we obtain the achievable uplink rate for the kth user as

$$R_{\mathrm{IP},k}^{\mathtt{zf}} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{p_{\mathrm{u}}}{\left(\sum_{i=1}^{K} \frac{p_{\mathrm{u}}\beta_i}{\tau p_{\mathrm{u}}\beta_i + 1} + 1 \right) \left[\left(\hat{\boldsymbol{G}}^H \hat{\boldsymbol{G}} \right)^{-1} \right]_{kk}} \right) \right\}. \tag{42}$$

Following the same derivations as in Section 3.1.2 for the case of perfect CSI, we obtain the following lower bound on the achievable uplink rate.

Proposition 7 *With ZF processing using imperfect CSI, Rayleigh fading, and for $M \geq K+1$, the achievable uplink rate for the kth user is bounded as*

$$\tilde{R}_{\mathrm{IP},k}^{\mathtt{zf}} = \log_2 \left(1 + \frac{\tau p_{\mathrm{u}}^2 (M-K) \beta_k^2}{(\tau p_{\mathrm{u}}\beta_k + 1) \sum_{i=1}^{K} \frac{p_{\mathrm{u}}\beta_i}{\tau p_{\mathrm{u}}\beta_i + 1} + \tau p_{\mathrm{u}}\beta_k + 1} \right). \tag{43}$$

Similarly, with $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$, when $M \to \infty$, the achievable uplink rate and its lower bound tend to the ones for MRC (see (41)), i.e.,

$$\tilde{R}_{\mathrm{IP},k}^{\mathtt{zf}} \to \log_2 \left(1 + \tau \beta_k^2 E_{\mathrm{u}}^2 \right), \; M \to \infty, \tag{44}$$

which equals the rate value obtained from Proposition 5.

3.2.3 MMSE Receiver

With imperfect CSI, the received vector at the BS can be rewritten as

$$\boldsymbol{y} = \sqrt{p_{\mathrm{u}}} \hat{\boldsymbol{G}} \boldsymbol{x} - \sqrt{p_{\mathrm{u}}} \boldsymbol{\mathcal{E}} \boldsymbol{x} + \boldsymbol{n}. \tag{45}$$

Therefore, for the MMSE receiver, the kth column of $\hat{\boldsymbol{A}}$ is given by

$$\hat{\boldsymbol{a}}_k = \left(\hat{\boldsymbol{G}} \hat{\boldsymbol{G}}^H + \frac{1}{p_{\mathrm{u}}} \mathtt{Cov}\left(-\sqrt{p_{\mathrm{u}}} \boldsymbol{\mathcal{E}} \boldsymbol{x} + \boldsymbol{n} \right) \right)^{-1} \hat{\boldsymbol{g}}_k = \frac{\hat{\boldsymbol{\Lambda}}_k^{-1} \hat{\boldsymbol{g}}_k}{\hat{\boldsymbol{g}}_k^H \hat{\boldsymbol{\Lambda}}_k^{-1} \hat{\boldsymbol{g}}_k + 1}, \tag{46}$$

where $\mathtt{Cov}(\boldsymbol{a})$ denotes the covariance matrix of a random vector $\boldsymbol{a}$, and

$$\hat{\boldsymbol{\Lambda}}_k \triangleq \sum_{i=1, i \neq k}^{K} \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + \left(\sum_{i=1}^{K} \frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i + 1} + \frac{1}{p_{\mathrm{u}}} \right) \boldsymbol{I}_M. \tag{47}$$

Similarly to in Remark 1, by using Cauchy-Schwarz' inequality, we can show that the MMSE receiver given by (46) is the optimal detector in the sense that it maximizes the rate given by (37).

Substituting (46) into (37), we get the achievable uplink rate for the kth user with MMSE receivers as

$$\begin{aligned} R_{\mathrm{P},k}^{\mathtt{mmse}} &= \mathbb{E}\left\{\log_2\left(1+\hat{\boldsymbol{g}}_k^H \hat{\boldsymbol{\Lambda}}_k^{-1}\hat{\boldsymbol{g}}_k\right)\right\} \\ &= \mathbb{E}\left\{\log_2\left(\frac{1}{\left[\left(\boldsymbol{I}_K+\left(\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i+1}+\frac{1}{p_{\mathrm{u}}}\right)^{-1}\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}}\right)^{-1}\right]_{kk}}\right)\right\}. \end{aligned} \tag{48}$$

Again, using an approximate distribution for the SINR, we can obtain a lower bound on the achievable uplink rate in closed form.

Proposition 8 *With imperfect CSI and Rayleigh fading, the achievable rate for the kth user with MMSE processing is approximately lower bounded as follows:*

$$\tilde{R}_{\mathrm{IP},k}^{\mathtt{mmse}} = \log_2\left(1+\left(\hat{\alpha}_k-1\right)\hat{\theta}_k\right), \tag{49}$$

where

$$\hat{\alpha}_k = \frac{\left(M-K+1+(K-1)\,\hat{\mu}\right)^2}{M-K+1+(K-1)\,\hat{\kappa}},\ \hat{\theta}_k = \frac{M-K+1+(K-1)\,\hat{\kappa}}{M-K+1+(K-1)\,\hat{\mu}}\omega\hat{\beta}_k, \tag{50}$$

where $\omega \triangleq \left(\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i+1}+\frac{1}{p_{\mathrm{u}}}\right)^{-1}$, $\hat{\beta}_k \triangleq \frac{\tau p_{\mathrm{u}}\beta_k^2}{\tau p_{\mathrm{u}}\beta_k+1}$, $\hat{\mu}$ *and* $\hat{\kappa}$ *are obtained by using following equations:*

$$\begin{aligned} \hat{\mu} &= \frac{1}{K-1}\sum_{i=1,i\neq k}^{K}\frac{1}{M\omega\hat{\beta}_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\hat{\mu}\right)+1} \\ \hat{\kappa}&\left(1+\sum_{i=1,i\neq k}^{K}\frac{\omega\hat{\beta}_i}{\left(M\omega\hat{\beta}_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\hat{\mu}\right)+1\right)^2}\right) \\ &= \sum_{i=1,i\neq k}^{K}\frac{\omega\hat{\beta}_i\hat{\mu}+1}{\left(M\omega\hat{\beta}_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\hat{\mu}\right)+1\right)^2}. \end{aligned} \tag{51}$$

Table 1 summarizes the lower bounds on the achievable rates for linear receivers derived in this section, distinguishing between the cases of perfect and imperfect CSI, respectively.

We have considered a *single-cell* MU-MIMO system. This simplifies the analysis, and it gives us important insights into how power can be scaled with the number of antennas in very large MIMO systems. A natural question is to what extent this power-scaling law still holds for *multicell* MU-MIMO systems. Intuitively, when we reduce the transmit power of each user, the effect of interference from other cells also reduces and hence, the SINR will stay unchanged. Therefore we will have the same power-scaling law as in the single-cell scenario. The next section explains this argument in more detail.

Table 1: Lower bounds on the achievable rates of the uplink transmission for the kth user.

	Perfect CSI	Imperfect CSI
MRC	$\log_2\left(1+\frac{p_{\mathrm{u}}(M-1)\beta_k}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}\beta_i+1}\right)$	$\log_2\left(1+\frac{\tau p_{\mathrm{u}}^2(M-1)\beta_k^2}{p_{\mathrm{u}}(\tau p_{\mathrm{u}}\beta_k+1)\sum_{i=1,i\neq k}^{K}\beta_i+(\tau+1)p_{\mathrm{u}}\beta_k+1}\right)$
ZF	$\log_2\left(1+p_{\mathrm{u}}\left(M-K\right)\beta_k\right)$	$\log_2\left(1+\frac{\tau p_{\mathrm{u}}^2(M-K)\beta_k^2}{(\tau p_{\mathrm{u}}\beta_k+1)\sum_{i=1}^{K}\frac{p_{\mathrm{u}}\beta_i}{\tau p_{\mathrm{u}}\beta_i+1}+\tau p_{\mathrm{u}}\beta_k+1}\right)$
MMSE	$\log_2\left(1+\left(\alpha_k-1\right)\theta_k\right)$	$\log_2\left(1+\left(\hat{\alpha}_k-1\right)\hat{\theta}_k\right)$

3.3 Power-Scaling Law for Multicell MU-MIMO Systems

We will use the MRC for our analysis. A similar analysis can be performed for the ZF and MMSE detectors. Consider the uplink of a multicell MU-MIMO system with L cells sharing the same frequency band. Each cell includes one BS equipped with M antennas and K single-antenna users. The $M\times 1$ received vector at the lth BS is given by

$$\boldsymbol{y}_l=\sqrt{p_{\mathrm{u}}}\sum_{i=1}^{L}\boldsymbol{G}_{li}\boldsymbol{x}_i+\boldsymbol{n}_l, \tag{52}$$

where $\sqrt{p_{\mathrm{u}}}\boldsymbol{x}_i$ is the $K\times 1$ transmitted vector of K users in the ith cell; $\boldsymbol{n}_l$ is an AWGN vector, $\boldsymbol{n}_l\sim\mathcal{CN}\left(0,\boldsymbol{I}_M\right)$; and $\boldsymbol{G}_{li}$ is the $M\times K$ channel matrix between the lth BS and the K users in the ith cell. The channel matrix $\boldsymbol{G}_{li}$ can be represented as

$$\boldsymbol{G}_{li}=\boldsymbol{H}_{li}\boldsymbol{D}_{li}^{1/2}, \tag{53}$$

where $\boldsymbol{H}_{li}$ is the fast fading matrix between the lth BS and the K users in the ith cell whose elements have zero mean and unit variance; and $\boldsymbol{D}_{li}$ is a $K\times K$ diagonal matrix, where $[\boldsymbol{D}_{li}]_{kk}=\beta_{lik}$, with β_{lik} represents the large-scale fading between the kth user in the i cell and the lth BS.

3.3.1 Perfect CSI

With perfect CSI, the received signal at the lth BS after using MRC is given by

$$\boldsymbol{r}_l = \boldsymbol{G}_{ll}^H \boldsymbol{y}_l = \sqrt{p_{\mathrm{u}}} \boldsymbol{G}_{ll}^H \boldsymbol{G}_{ll} \boldsymbol{x}_l + \sqrt{p_{\mathrm{u}}} \sum_{i=1, i \neq l}^{L} \boldsymbol{G}_{ll}^H \boldsymbol{G}_{li} \boldsymbol{x}_i + \boldsymbol{G}_{ll}^H \boldsymbol{n}_l. \tag{54}$$

With $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{M}$, (54) can be rewritten as

$$\frac{1}{\sqrt{M}} \boldsymbol{r}_l = \sqrt{E_{\mathrm{u}}} \frac{\boldsymbol{G}_{ll}^H \boldsymbol{G}_{ll}}{M} \boldsymbol{x}_l + \sqrt{p_{\mathrm{u}}} \sum_{i=1, i \neq l}^{L} \frac{\boldsymbol{G}_{ll}^H \boldsymbol{G}_{li}}{M} \boldsymbol{x}_i + \frac{1}{\sqrt{M}} \boldsymbol{G}_{ll}^H \boldsymbol{n}_l. \tag{55}$$

From (5)–(6), when M grows large, the interference from other cells disappears. More precisely,

$$\frac{1}{\sqrt{M}} \boldsymbol{r}_l \to \sqrt{E_{\mathrm{u}}} \boldsymbol{D}_{ll} \boldsymbol{x}_l + \boldsymbol{D}_{ll}^{1/2} \tilde{\boldsymbol{n}}_l, \tag{56}$$

where $\tilde{\boldsymbol{n}}_l \sim \mathcal{CN}(0, \boldsymbol{I})$. Therefore, the SINR of the uplink transmission from the kth user in the lth cell converges to a constant value when M grows large, more precisely

$$\mathsf{SINR}_{l,k}^{\mathrm{P}} \to \beta_{llk} E_{\mathrm{u}}, \text{ as } M \to \infty. \tag{57}$$

This means that the power scaling law derived for single-cell systems is valid in multicell systems too.

3.3.2 Imperfect CSI

In this case, the channel estimate from the uplink pilots is contaminated by interference from other cells. The MMSE channel estimate of the channel matrix $\boldsymbol{G}_{ll}$ is given by [10]

$$\hat{\boldsymbol{G}}_{ll} = \left(\sum_{i=1}^{L} \boldsymbol{G}_{li} + \frac{1}{\sqrt{p_{\mathrm{p}}}} \boldsymbol{W}_l \right) \tilde{\boldsymbol{D}}_{ll}, \tag{58}$$

where $\tilde{\boldsymbol{D}}_{ll}$ is a diagonal matrix where the kth diagonal element $\left[\tilde{\boldsymbol{D}}_{ll}\right]_{kk} = \beta_{llk} \left(\sum_{i=1}^{L} \beta_{lik} + \frac{1}{p_{\mathrm{p}}} \right)^{-1}$. The received signal at the lth BS after using MRC is given by

$$\hat{\boldsymbol{r}}_l = \hat{\boldsymbol{G}}_{ll}^H \boldsymbol{y}_l = \tilde{\boldsymbol{D}}_{ll} \left(\sum_{i=1}^{L} \boldsymbol{G}_{li} + \frac{1}{\sqrt{p_{\mathrm{p}}}} \boldsymbol{W}_l \right)^H \left(\sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \boldsymbol{G}_{li} \boldsymbol{x}_i + \boldsymbol{n}_l \right). \tag{59}$$

With $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$, we have

$$\frac{1}{M^{3/4}}\tilde{\boldsymbol{D}}_{ll}^{-1}\hat{\boldsymbol{r}}_l = \sqrt{E_{\mathrm{u}}}\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\boldsymbol{G}_{li}^H\boldsymbol{G}_{lj}}{M}\boldsymbol{x}_j + \sum_{i=1}^{L}\frac{\boldsymbol{G}_{li}^H\boldsymbol{n}_l}{M^{3/4}} + \frac{1}{\sqrt{\tau}}\sum_{i=1}^{L}\frac{\boldsymbol{W}_l^H\boldsymbol{G}_{li}}{M^{3/4}}\boldsymbol{x}_i + \frac{1}{\sqrt{\tau E_{\mathrm{u}}}}\frac{\boldsymbol{W}_l^H\boldsymbol{n}_l}{M^{1/2}}. \tag{60}$$

By using (5) and (6), as M grows large, we obtain

$$\frac{1}{M^{3/4}}\tilde{\boldsymbol{D}}_{ll}^{-1}\hat{\boldsymbol{r}}_l \to \sqrt{E_{\mathrm{u}}}\sum_{i=1}^{L}\boldsymbol{D}_{li}\boldsymbol{x}_i + \frac{1}{\sqrt{\tau E_{\mathrm{u}}}}\tilde{\boldsymbol{w}}_l, \tag{61}$$

where $\tilde{\boldsymbol{w}}_l \sim \mathcal{CN}(0, \boldsymbol{I}_M)$. Therefore, the asymptotic SINR of the uplink from the kth user in the lth cell is

$$\mathsf{SINR}_{l,k}^{\mathrm{IP}} \to \frac{\tau\beta_{llk}^2 E_{\mathrm{u}}^2}{\tau\sum_{i\neq l}^{L}\beta_{lik}^2 E_{\mathrm{u}}^2 + 1}, \text{ as } M \to \infty. \tag{62}$$

We can see that the $1/\sqrt{M}$ power-scaling law still holds. Furthermore, transmission from users in other cells constitutes residual interference. The reason is that the pilot reuse gives pilot-contamination-induced inter-cell interference which grows with M at the same rate as the desired signal.

Remark 5 *The MMSE channel estimate* (10) *is obtained by the assumption that, for uplink training, all cells simultaneously transmit pilot sequences, and that the same set of pilot sequences is used in all cells. This assumption makes no fundamental difference compared with using different pilot sequences in different cells, as explained [7, Section VII-F]. Nor does this assumption make any fundamental difference to the case when users in other cells transmit* data *when the users in the cell of interest send their pilots. The reason is that whatever data is transmitted in other cells, it can always be expanded in terms of the orthogonal pilot sequences that are transmitted in the cell of interest, so pilot contamination ensues. For example, consider the uplink training in cell* 1 *of a MU-MIMO system with* $L = 2$ *cells. Assume that, during an interval of length* τ *symbols* ($\tau \geq K$)*,* K *users in cell* 1 *are transmitting uplink pilots* $\boldsymbol{\Phi}^T$ *at the same time as* K *users in cell* 2 *are transmitting uplink data* $\boldsymbol{X}_2$*. Here* $\boldsymbol{\Phi}$ *is a* $\tau \times K$ *matrix which satisfies* $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{I}_K$*. The received signal at base station* 1 *is*

$$\boldsymbol{Y}_1 = \sqrt{p_{\mathrm{p}}}\boldsymbol{G}_{11}\boldsymbol{\Phi}^T + \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\boldsymbol{X}_2 + \boldsymbol{N}_1,$$

where $\boldsymbol{N}_1 \in \mathbb{C}^{M\times\tau}$ *is AWGN at base station* 1*. By projecting the received signal* $\boldsymbol{Y}_1$ *onto* $\boldsymbol{\Phi}^*$*, we obtain*

$$\tilde{\boldsymbol{Y}}_1 \triangleq \boldsymbol{Y}_1\boldsymbol{\Phi}^* = \sqrt{p_{\mathrm{p}}}\boldsymbol{G}_{11} + \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\tilde{\boldsymbol{X}}_2 + \tilde{\boldsymbol{N}}_1,$$

where $\tilde{\boldsymbol{X}}_2 \triangleq \boldsymbol{X}_2\boldsymbol{\Phi}^*$, *and* $\tilde{\boldsymbol{N}}_1 \triangleq \boldsymbol{N}_1\boldsymbol{\Phi}^*$. *The* k*th column of* $\tilde{\boldsymbol{Y}}_1$ *is given by*

$$\tilde{\boldsymbol{y}}_{1k} = \sqrt{p_{\mathrm{p}}}\boldsymbol{g}_{11k} + \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\tilde{\boldsymbol{x}}_{2k} + \tilde{\boldsymbol{n}}_{1k},$$

where $\boldsymbol{g}_{11k}$, $\tilde{\boldsymbol{x}}_{2k}$, *and* $\tilde{\boldsymbol{n}}_{1k}$ *are the* k*th columns of* $\boldsymbol{G}_{11}$, $\tilde{\boldsymbol{X}}_2$, *and* $\tilde{\boldsymbol{N}}_1$, *respectively. By using the Lindeberg-Lévy central limit theorem, we find that each element of the vector* $\sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\tilde{\boldsymbol{x}}_{2,k}$ *(ignoring the large-scale fading in this argument) is approximately Gaussian distributed with zero mean and variance* Kp_{u}. *If* $K = \tau$, *then* $Kp_{\mathrm{u}} = p_{\mathrm{p}}$ *and this result means that the effect of payload interference is just as bad as if users in cell* 2 *transmitted pilot sequences.*

4 Energy-Efficiency versus Spectral-Efficiency Tradeoff

The energy-efficiency (in bits/Joule) of a system is defined as the spectral-efficiency (sum-rate in bits/channel use) divided by the transmit power expended (in Joules/channel use). Typically, increasing the spectral efficiency is associated with increasing the power and hence, with decreasing the energy-efficiency. Therefore, there is a fundamental tradeoff between the energy efficiency and the spectral efficiency. However, in one operating regime it is possible to jointly increase the energy and spectral efficiencies, and in this regime there is no tradeoff. This may appear a bit counterintuitive at first, but it falls out from the analysis in Section 4.1. Note, however, that this effect occurs in an operating regime that is probably of less interest in practice.

In this section, we study the energy-spectral efficiency tradeoff for the uplink of MU-MIMO systems using linear receivers at the BS. Certain activities (multiplexing to many users rather than beamforming to a single user and increasing the number of service antennas) can simultaneously benefit both the spectral-efficiency and the radiated energy-efficiency. Once the number of service antennas is set, one can adjust other system parameters (radiated power, numbers of users, duration of pilot sequences) to obtain increased spectral-efficiency at the cost of reduced energy-efficiency, and vice-versa. This should be a desirable feature for service providers: they can set the operating point according to the current traffic demand (high energy-efficiency and low spectral-efficiency, for example, during periods of low demand).

4.1 Single-Cell MU-MIMO Systems

We define the spectral efficiency for perfect and imperfect CSI, respectively, as follows

$$R_{\mathrm{P}}^{A} = \sum_{k=1}^{K} \tilde{R}_{\mathrm{P},k}^{A}, \text{ and } R_{\mathrm{IP}}^{A} = \frac{T-\tau}{T} \sum_{k=1}^{K} \tilde{R}_{\mathrm{IP},k}^{A}, \tag{63}$$

where $A \in \{\texttt{mrc}, \texttt{zf}, \texttt{mmse}\}$ corresponds to MRC, ZF and MMSE, and T is the coherence interval in symbols. The energy-efficiency for perfect and imperfect CSI is defined as

$$\eta_{\mathrm{P}}^{A} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{P}}^{A}, \text{ and } \eta_{\mathrm{IP}}^{A} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{A}. \tag{64}$$

The large-scale fading can be incorporated by substituting (40) and (43) into (63). However, this yields energy and spectral efficiency formulas of an intractable form and which are very difficult (if not impossible) to use for obtaining further insights. Note that the large number of antennas effectively removes the small-scale fading, but the effects of path loss and large-scale fading will remain. This may give different users vastly different SNRs. As a result, power control may be desired. In principle, a power control factor could be included by letting p_{u} in (40) and (43) depend on k. The optimal transmit power for each user would depend only on the large-scale fading, not on the small-scale fading and effective power-control rules could be developed straightforwardly from the resulting expressions. However, the introduction of such power control may bring new trade-offs, for example that of fairness between users near and far from the BS. In addition, the spectral versus energy efficiency tradeoff relies on optimization of the number of active users. If the users have grossly different large-scale fading coefficients, then the issue will arise as to whether these coefficients should be fixed before the optimization or whether for a given number of users K, these coefficients should be drawn randomly. Both ways can be justified, but have different operational meaning in terms of scheduling. This leads, among others, to issues with fairness versus total throughput, which we would like to avoid here as this matter could easily obscure the main points of our analysis. Therefore, for analytical tractability, we ignore the effect of the large-scale fading here, i.e., we set $\boldsymbol{D} = \boldsymbol{I}_K$. Also, we only consider MRC and ZF receivers.[3]

For perfect CSI, it is straightforward to show from (17), (21), and (64) that when the spectral efficiency increases, the energy efficiency decreases. For imperfect CSI, this is not always so, as we shall see next. In what follows, we focus on the case of imperfect CSI since this is the case of interest in practice.

4.1.1 Maximum-Ratio Combining

From (40), the spectral efficiency and energy efficiency with MRC processing are given by

$$R_{\mathrm{IP}}^{\texttt{mrc}} = \frac{T-\tau}{T} K \log_2 \left(1 + \frac{\tau\,(M-1)\,p_{\mathrm{u}}^2}{\tau\,(K-1)\,p_{\mathrm{u}}^2 + (K+\tau)\,p_{\mathrm{u}} + 1}\right), \text{ and}$$
$$\eta_{\mathrm{IP}}^{\texttt{mrc}} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{\texttt{mrc}}. \tag{65}$$

[3]When M is large, the performance of the MMSE receiver is very close to that of the ZF receiver (see Section 5). Therefore, the insights on energy versus spectral efficiency obtained from studying the performance of ZF can be used to draw conclusions about MMSE as well.

We have

$$\begin{aligned}\lim_{p_{\mathrm{u}}\to 0}\eta_{\mathrm{IP}}^{\mathtt{mrc}} &= \lim_{p_{\mathrm{u}}\to 0}\frac{1}{p_{\mathrm{u}}}R_{\mathrm{IP}}^{\mathtt{mrc}} \\ &= \lim_{p_{\mathrm{u}}\to 0}\frac{T-\tau}{T}K\frac{(\log_2 e)\,\tau\,(M-1)\,p_{\mathrm{u}}}{\tau\,(K-1)\,p_{\mathrm{u}}^2+(K+\tau)\,p_{\mathrm{u}}+1}=0,\end{aligned} \tag{66}$$

and

$$\lim_{p_{\mathrm{u}}\to\infty}\eta_{\mathrm{IP}}^{\mathtt{mrc}} = \lim_{p_{\mathrm{u}}\to\infty}\frac{1}{p_{\mathrm{u}}}R_{\mathrm{IP}}^{\mathtt{mrc}} = 0. \tag{67}$$

Equations (66) and (67) imply that for low p_{u}, the energy efficiency increases when p_{u} increases, and for high p_{u} the energy efficiency decreases when p_{u} increases. Since $\frac{\partial R_{\mathrm{IP}}^{\mathtt{mrc}}}{\partial p_{\mathrm{u}}} > 0$, $\forall p_{\mathrm{u}} > 0$, $R_{\mathrm{IP}}^{\mathtt{mrc}}$ is a monotonically increasing function of p_{u}. Therefore, at low p_{u} (and hence at low spectral efficiency), the energy efficiency increases as the spectral efficiency increases and vice versa at high p_{u}. The reason is that, the spectral efficiency suffers from a "squaring effect" when the received data signal is multiplied with the received pilots. Hence, at $p_{\mathrm{u}} \ll 1$, the spectral-efficiency behaves as $\sim p_{\mathrm{u}}^2$. As a consequence, the energy efficiency (which is defined as the spectral efficiency divided by p_{u}) increases linearly with p_{u}. In more detail, expanding the rate in a Taylor series for $p_{\mathrm{u}} \ll 1$, we obtain

$$\begin{aligned}R_{\mathrm{IP}}^{\mathtt{mrc}} &\approx R_{\mathrm{IP}}^{\mathtt{mrc}}|_{p_{\mathrm{u}}=0} + \left.\frac{\partial R_{\mathrm{IP}}^{\mathtt{mrc}}}{\partial p_{\mathrm{u}}}\right|_{p_{\mathrm{u}}=0} p_{\mathrm{u}} + \frac{1}{2}\left.\frac{\partial^2 R_{\mathrm{IP}}^{\mathtt{mrc}}}{\partial p_{\mathrm{u}}^2}\right|_{p_{\mathrm{u}}=0} p_{\mathrm{u}}^2 \\ &= \frac{T-\tau}{T}K\log_2(e)\,\tau\,(M-1)\,p_{\mathrm{u}}^2.\end{aligned} \tag{68}$$

This gives the following relation between the spectral efficiency and energy efficiency at $p_{\mathrm{u}} \ll 1$:

$$\eta_{\mathrm{IP}}^{\mathtt{mrc}} = \sqrt{\frac{T-\tau}{T}K\log_2(e)\,\tau\,(M-1)\,R_{\mathrm{IP}}^{\mathtt{mrc}}}. \tag{69}$$

We can see that when $p_{\mathrm{u}} \ll 1$, by doubling the spectral efficiency, or by doubling M, we can increase the energy efficiency by 1.5 dB.

4.1.2 Zero-Forcing Receiver

From (43), the spectral efficiency and energy efficiency for ZF are given by

$$R_{\mathrm{IP}}^{\mathtt{zf}} = \frac{T-\tau}{T}K\log_2\left(1+\frac{\tau\,(M-K)\,p_{\mathrm{u}}^2}{(K+\tau)\,p_{\mathrm{u}}+1}\right),\ \text{and}\ \eta_{\mathrm{IP}}^{\mathtt{zf}} = \frac{1}{p_{\mathrm{u}}}R_{\mathrm{IP}}^{\mathtt{zf}}. \tag{70}$$

Similarly to in the analysis of MRC, we can show that at low transmit power p_{u}, the energy efficiency increases when the spectral efficiency increases. In the low-p_{u} regime, we obtain the following Taylor series expansion

$$R_{\mathrm{IP}}^{\mathtt{zf}} \approx \frac{T-\tau}{T} K \log_2(e)\, \tau\,(M-K)\, p_{\mathrm{u}}^2, \text{ for } p_{\mathrm{u}} \ll 1. \tag{71}$$

Therefore,

$$\eta_{\mathrm{IP}}^{\mathtt{zf}} = \sqrt{\frac{T-\tau}{T} K \log_2(e)\, \tau\,(M-K)\, R_{\mathrm{IP}}^{\mathtt{zf}}}. \tag{72}$$

Again, at $p_{\mathrm{u}} \ll 1$, by doubling M or $R_{\mathrm{IP}}^{\mathtt{zf}}$, we can increase the energy efficiency by 1.5 dB.

4.2 Multicell MU-MIMO Systems

In this section, we derive expressions for the energy-efficiency and spectral-efficiency for a multicell system. These are used for the simulation in the Section 5. Here, we consider a simplified channel model, i.e., $\boldsymbol{D}_{ll} = \boldsymbol{I}_K$, and $\boldsymbol{D}_{li} = \beta \boldsymbol{I}_K$, where $\beta \in [0,1]$ is an intercell interference factor. Note that from (10), the estimate of the channel between the kth user in the lth cell and the lth BS is given by

$$\hat{\boldsymbol{g}}_{llk} = \left((L-1)\,\beta + 1 + \frac{1}{p_{\mathrm{p}}}\right)^{-1} \left(\boldsymbol{h}_{llk} + \sum_{i \neq k}^{L} \sqrt{\beta}\boldsymbol{h}_{lik} + \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{w}_{lk}\right). \tag{73}$$

The term $\sum_{i\neq k}^{L} \sqrt{\beta}\boldsymbol{h}_{lik}$ represents the pilot contamination, therefore

$$\frac{\sum_{i\neq k}^{L} \mathbb{E}\left\{\|\sqrt{\beta}\boldsymbol{h}_{lik}\|^2\right\}}{\mathbb{E}\left\{\|\boldsymbol{h}_{llk}\|^2\right\}} = \beta\,(L-1)$$

can be considered as the effect of pilot contamination.

Following a similar derivation as in the case of single-cell MU-MIMO systems, we obtain the spectral efficiency and energy efficiency for imperfect CSI with MRC and ZF receivers, respectively, as follows:

$$R_{\mathrm{mul}}^{\mathtt{mrc}} = \frac{T-\tau}{T} K \log_2\left(1 + \frac{\tau\,(M-1)\,p_{\mathrm{u}}^2}{\tau\left(K\bar{L}^2 - 1 + \beta\left(\bar{L}-1\right)(M-2)\right)p_{\mathrm{u}}^2 + \bar{L}\,(K+\tau)\,p_{\mathrm{u}} + 1}\right),$$

$$\eta_{\mathrm{mul}}^{\mathtt{mrc}} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{\mathtt{mrc}} \tag{74}$$

$$R_{\mathrm{mul}}^{\mathtt{zf}} = \frac{T-\tau}{T} K \log_2\left(1 + \frac{\tau\,(M-K)\,p_{\mathrm{u}}^2}{\tau K\left(\bar{L}^2 - \bar{L}\beta + \beta - 1\right)p_{\mathrm{u}}^2 + \bar{L}\,(K+\tau)\,p_{\mathrm{u}} + 1}\right),$$

$$\eta_{\mathrm{IP}}^{\mathtt{zf}} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{ml}}^{\mathtt{zf}}, \tag{75}$$

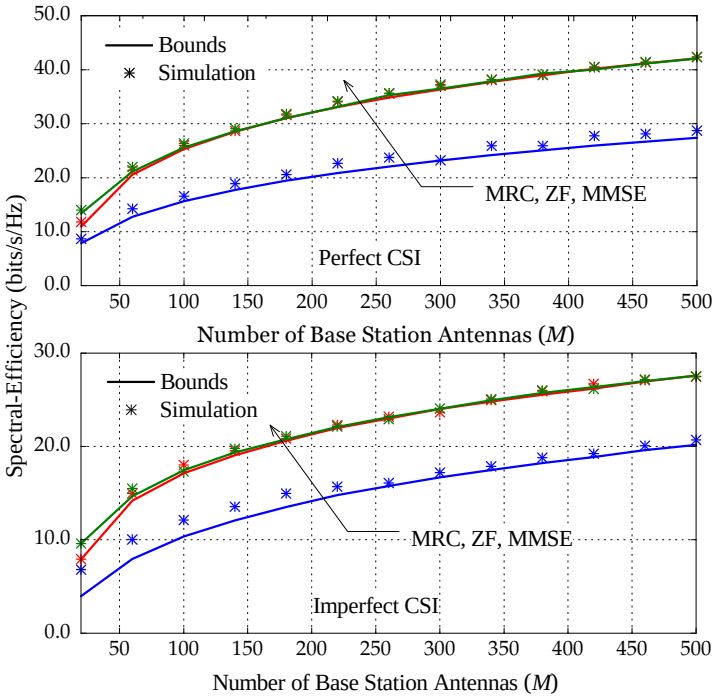


Figure 1: Lower bounds and numerically evaluated values of the spectral efficiency for different numbers of BS antennas for MRC, ZF, and MMSE with perfect and imperfect CSI. In this example there are $K = 10$ users, the coherence interval $T = 196$, the transmit power per terminal is $p_{\mathrm{u}} = 10$ dB, and the propagation channel parameters were $\sigma_{\mathrm{shadow}} = 8$ dB, and $\nu = 3.8$.

where $\bar{L} \triangleq (L-1)\beta+1$. The principal complexity in the derivation is the correlation between pilot-contaminated channel estimates.

We can see that the spectral efficiency is a decreasing function of β and L. Furthermore, when $L = 1$, or $\beta = 0$, the results (74) and (75) coincide with (65) and (70) for single-cell MU-MIMO systems.

5 Numerical Results

5.1 Single-Cell MU-MIMO Systems

We consider a hexagonal cell with a radius (from center to vertex) of 1000 meters. The users are located uniformly at random in the cell and we assume that no user

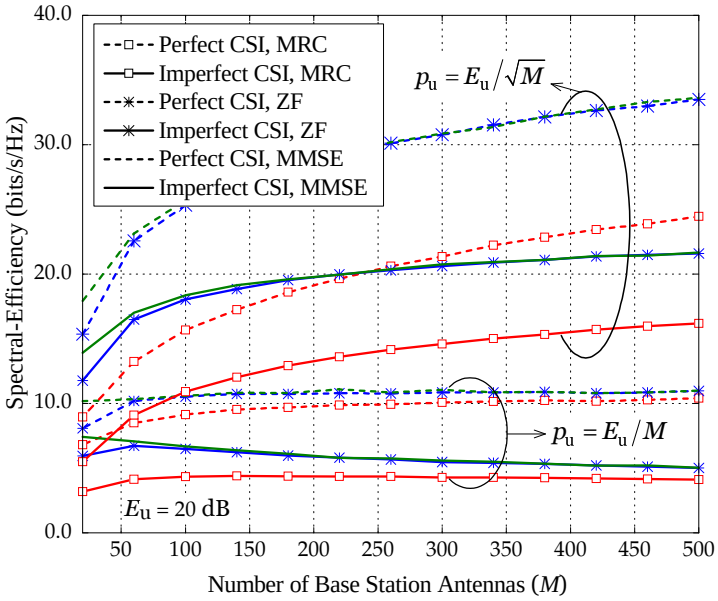


Figure 2: Spectral efficiency versus the number of BS antennas M for MRC, ZF, and MMSE processing at the receiver, with perfect CSI and with imperfect CSI (obtained from uplink pilots). In this example $K = 10$ users are served simultaneously, the reference transmit power is $E_u = 20$ dB, and the propagation parameters were $\sigma_{\text{shadow}} = 8$ dB and $\nu = 3.8$.

is closer to the BS than $r_h = 100$ meters. The large-scale fading is modelled via $\beta_k = z_k/(r_k/r_h)^\nu$, where z_k is a log-normal random variable with standard deviation σ_{shadow}, r_k is the distance between the kth user and the BS, and ν is the path loss exponent. For all examples, we choose $\sigma_{\text{shadow}} = 8$ dB, and $\nu = 3.8$.

We assume that the transmitted data are modulated with OFDM. Here, we choose parameters that resemble those of LTE standard: an OFDM symbol duration of $T_s = 71.4\mu$s, and a useful symbol duration of $T_u = 66.7\mu$s. Therefore, the guard interval length is $T_g = T_s - T_u = 4.7\mu$s. We choose the channel coherence time to be $T_c = 1$ ms. Then, $T = \frac{T_c}{T_s}\frac{T_u}{T_g} = 196$, where $\frac{T_c}{T_s} = 14$ is the number of OFDM symbols in a 1 ms coherence interval, and $\frac{T_u}{T_g} = 14$ corresponds to the "frequency smoothness interval" [7].

5.1.1 Power-Scaling Law

We first conduct an experiment to validate the tightness of our proposed capacity bounds. Fig. 1 shows the simulated spectral efficiency and the proposed analytical

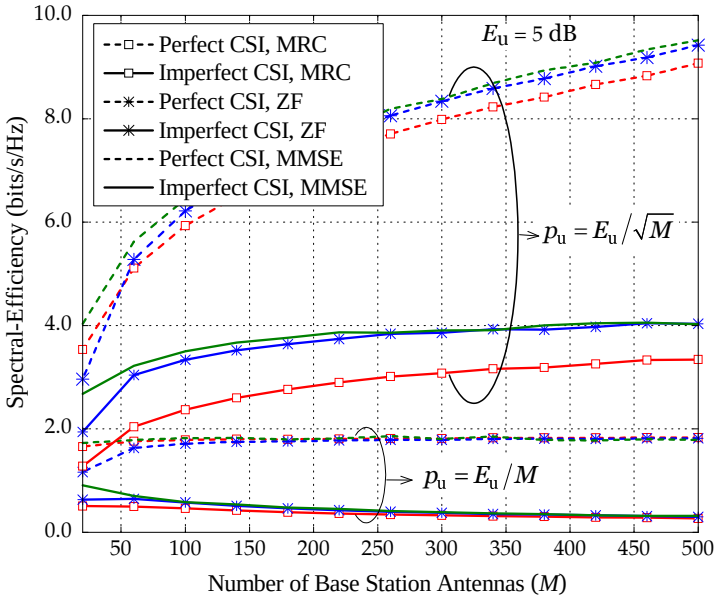


Figure 3: Same as Figure 2, but with $E_u = 5$ dB.

bounds for MRC, ZF, and MMSE receivers with perfect and imperfect CSI at $p_u = 10$ dB. In this example there are $K = 10$ users. For CSI estimation from uplink pilots, we choose pilot sequences of length $\tau = K$. (This is the smallest amount of training that can be used.) Clearly, all bounds are very tight, especially at large M. Therefore, in the following, we will use these bounds for all numerical work.

We next illustrate the power scaling laws. Fig. 2 shows the spectral efficiency on the uplink versus the number of BS antennas for $p_u = E_u/M$ and $p_u = E_u/\sqrt{M}$ with perfect and imperfect receiver CSI, and with MRC, ZF, and MMSE processing, respectively. Here, we choose $E_u = 20$ dB. At this SNR, the spectral efficiency is in the order of 10–30 bits/s/Hz, corresponding to a spectral efficiency per user of 1–3 bits/s/Hz. These operating points are reasonable from a practical point of view. For example, 64-QAM with a rate-1/2 channel code would correspond to 3 bits/s/Hz. (Figure 3, see below, shows results at lower SNR.) As expected, with $p_u = E_u/M$, when M increases, the spectral efficiency approaches a constant value for the case of perfect CSI, but decreases to 0 for the case of imperfect CSI. However, with $p_u = E_u/\sqrt{M}$, for the case of perfect CSI the spectral efficiency grows without bound (logarithmically fast with M) when $M \to \infty$ and with imperfect CSI, the spectral efficiency converges to a nonzero limit as $M \to \infty$. These results confirm that we can scale down the transmitted power of each user as E_u/M for the perfect CSI case, and as $E_u/\sqrt{M}$ for the imperfect CSI case when M is large.

Typically ZF is better than MRC at high SNR, and vice versa at low SNR [12]. MMSE always performs the best across the entire SNR range (see Remark 1). When

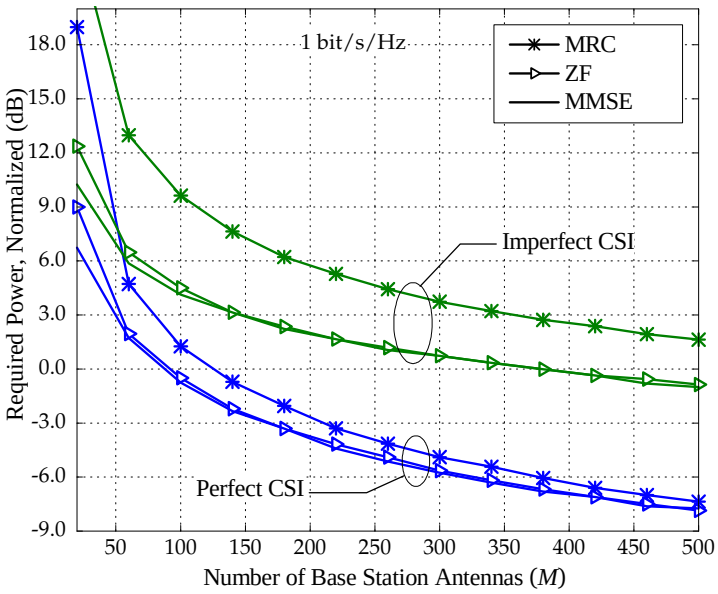


Figure 4: Transmit power required to achieve 1 bit/channel use per user for MRC, ZF, and MMSE processing, with perfect and imperfect CSI, as a function of the number M of BS antennas. The number of users is fixed to $K = 10$, and the propagation parameters are $\sigma_{\text{shadow}} = 8$ dB and $\nu = 3.8$.

comparing MRC and ZF in Fig. 2, we see that here, when the transmitted power is proportional to $1/\sqrt{M}$, the power is not low enough to make MRC perform as well as ZF. But when the transmitted power is proportional to $1/M$, MRC performs almost as well as ZF for large M. Furthermore, as we can see from the figure, MMSE is always better than MRC or ZF, and its performance is very close to ZF.

In Fig. 3, we consider the same setting as in Fig. 2, but we choose $E_{\text{u}} = 5$ dB. This figure provides the same insights as Fig. 2. The gap between the performance of MRC and that of ZF (or MMSE) is reduced compared with Fig. 2. This is so because the relative effect of crosstalk interference (the interference from other users) as compared to the thermal noise is smaller here than in Fig. 2.

We next show the transmit power per user that is needed to reach a fixed spectral efficiency. Fig. 4 shows the normalized power (p_{u}) required to achieve 1 bit/s/Hz per user as a function of M. As predicted by the analysis, by doubling M, we can cut back the power by approximately 3 dB and 1.5 dB for the cases of perfect and imperfect CSI, respectively. When M is large ($M/K \gtrapprox 6$), the difference in performance between MRC and ZF (or MMSE) is less than 1 dB and 3 dB for the cases of perfect and imperfect CSI, respectively. This difference increases when we increase the target spectral efficiency. Fig. 2 shows the normalized power required for 2 bit/s/Hz per user. Here, the crosstalk interference is more significant (relative

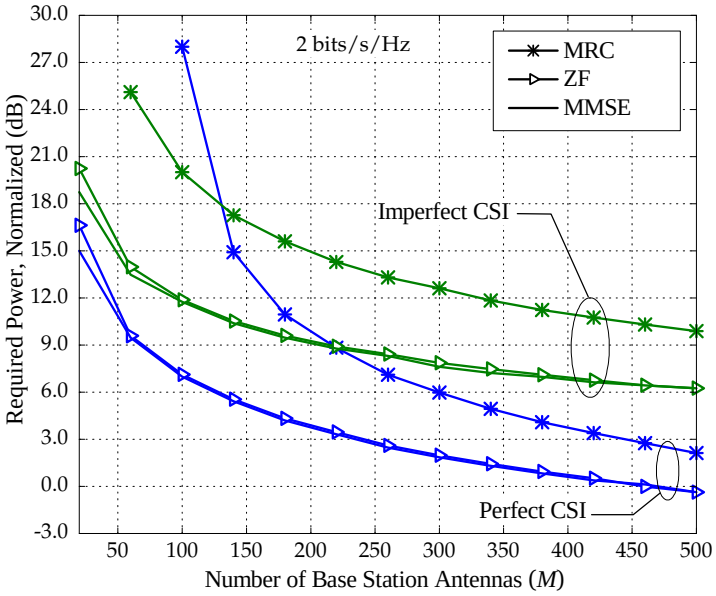


Figure 5: Same as Figure 4 but for a target spectral efficiency of 2 bits/channel use per user.

to the thermal noise) and hence the ZF and MMSE receivers perform relatively better.

5.1.2 Energy Efficiency versus Spectral Efficiency Tradeoff

We next examine the tradeoff between energy efficiency and spectral efficiency in more detail. Here, we ignore the effect of large-scale fading, i.e., we set $\boldsymbol{D} = \boldsymbol{I}_K$. We normalize the energy efficiency against a reference mode corresponding to a single-antenna BS serving one single-antenna user with $p_{\mathrm{u}} = 10$ dB. For this reference mode, the spectral efficiencies and energy efficiencies for MRC, ZF, and MMSE are equal, and given by (from (39) and (63))

$$R_{\mathrm{IP}}^{0} = \frac{T-\tau}{T}\,\mathbb{E}\left\{\log_2\left(1+\frac{\tau p_{\mathrm{u}}^2|z|^2}{1+p_{\mathrm{u}}\,(1+\tau)}\right)\right\},\ \eta_{\mathrm{IP}}^{0} = R_{\mathrm{IP}}^{0}/p_{\mathrm{u}},$$

where z is a Gaussian RV with zero mean and unit variance. For the reference mode, the spectral-efficiency is obtained by choosing the duration of the uplink pilot sequence τ to maximize R_{IP}^{0}. Numerically we find that $R_{\mathrm{IP}}^{0} = 2.65$ bits/s/Hz and $\eta_{\mathrm{IP}}^{0} = 0.265$ bits/J.

Fig. 5 shows the relative energy efficiency versus the the spectral efficiency for MRC and ZF. The relative energy efficiency is obtained by normalizing the energy

efficiency by η_{IP}^{0} and it is therefore dimensionless. The dotted and dashed lines show the performances for the cases of $M = 1, K = 1$ and $M = 100, K = 1$, respectively. Each point on the curves is obtained by choosing the transmit power p_{u} and pilot sequence length τ to maximize the energy efficiency for a given spectral efficiency. The solid lines show the performance for the cases of $M = 50$, and 100. Each point on these curves is computed by jointly choosing K, τ, and p_{u} to maximize the energy-efficiency subject a fixed spectral-efficiency, i.e.,

$$\arg \max_{p_{\mathrm{u}},K,\tau} \eta_{\mathrm{IP}}^{\mathtt{A}}, \qquad \text{s.t. } R_{\mathrm{IP}}^{\mathtt{A}} = \text{const.}, K \leq \tau \leq T.$$

We first consider a single-user system with $K = 1$. We compare the performance of the cases $M = 1$ and $M = 100$. Since $K = 1$ the performances of MRC and ZF are equal. With the same power used as in the reference mode, i.e., $p_{\mathrm{u}} = 10$ dB, using 100 antennas can increase the spectral efficiency and the energy efficiency by factors of 4 and 3, respectively. Reducing the transmit power by a factor of 100, from 10 dB to -10 dB yields a 100-fold improvement in energy efficiency compared with that of the reference mode with no reduction in spectral-efficiency.

We next consider a multiuser system ($K > 1$). Here the transmit power p_{u}, the number of users K, and the duration of pilot sequences τ are chosen optimally for fixed M. We consider $M = 50$ and 100. Here the system performance improves very significantly compared to the single-user case. For example, with MRC, at $p_{\mathrm{u}} = 0$ dB, compared with the case of $M = 1, K = 1$, the spectral-efficiency increases by factors of 50 and 80, while the energy-efficiency increases by factors of 55 and 75 for $M = 50$ and $M = 100$, respectively. As discussed in Section 4, at low spectral efficiency, the energy efficiency increases when the spectral efficiency increases. Furthermore, we can see that at high spectral efficiency, ZF outperforms MRC. This is due to the fact that the MRC receiver is limited by the intracell interference, which is significant at high spectral efficiency. As a consequence, when p_{u} is increased, the spectral efficiency of MRC approaches a constant value, while the energy efficiency goes to zero (see (67)).

The corresponding optimum values of K and τ as functions of the spectral efficiency for $M = 100$ are shown in Fig. 6. For MRC, the optimal number of users and uplink pilots are the same (this means that the minimal possible length of training sequences are used). For ZF, more of the coherence interval is used for training. Generally, at low transmit power and therefore at low spectral efficiency, we spend more time on training than on payload data transmission. At high power (high spectral efficiency and low energy efficiency), we can serve around 55 users, and $K = \tau$ for both MRC and ZF.

5.2 Multicell MU-MIMO Systems

Next, we examine the effect of pilot contamination on the energy and spectral efficiency for multicell systems. We consider a system with $L = 7$ cells. Each

cell has the same size as in the single-cell system. When shrinking the cell size, one typically also cuts back on the power. Hence, the relation between signal and interference power would not be substantially different in systems with smaller cells and in that sense, the analysis is largely independent of the actual physical size of the cell [21]. Note that, setting $L = 7$ means that we consider the performance of a given cell with the interference from 6 nearest-neighbor cells. We assume $\boldsymbol{D}_{ll} = \boldsymbol{I}_K$, and $\boldsymbol{D}_{li} = \beta \boldsymbol{I}_K$, for $i \neq l$. To examine the performance in a practical scenario, the intercell interference factor, β, is chosen as follows. We consider two users, the 1st user is located uniformly at random in the first cell, and the 2nd user is located uniformly at random in one of the 6 nearest-neighbor cells of the 1st cell. Let $\bar{\beta}_1$ and $\bar{\beta}_2$ be the large scale fading from the 1st user and the 2nd user to the 1st BS, respectively. (The large scale fading is modelled as in Section 5.1.1.) Then we compute β as $\mathbb{E}\left\{\bar{\beta}_2/\bar{\beta}_1\right\}$. By simulation, we obtain $\beta = 0.32, 0.11$, and 0.04 for the cases of ($\sigma_{\text{shadow}} = 8$ dB, $\nu = 3.8, f_{\text{reuse}} = 1$), ($\sigma_{\text{shadow}} = 8$ dB, $\nu = 3, f_{\text{reuse}} = 1$), and ($\sigma_{\text{shadow}} = 8$ dB, $\nu = 3.8, f_{\text{reuse}} = 3$), respectively, where f_{reuse} is the frequency reuse factor.

Fig. 7 shows the relative energy efficiency versus the spectral efficiency for MRC and ZF of the multicell system. The reference mode is the same as the one in Fig. 5 for a single-cell system. The dotted line shows the performance for the case of $M = 1, K = 1$, and $\beta = 0$. The solid and dashed lines show the performance for the cases of $M = 100$, and $L = 7$, with different intercell interference factors β of $0.32, 0.11$, and 0.04. Each point on these curves is computed by jointly choosing τ, K, and p_{u} to maximize the energy efficiency for a given spectral efficiency. We can see that the pilot contamination significantly degrades the system performance. For example, when β increases from 0.11 to 0.32 (and hence, the pilot contamination increases), with the same power, $p_{\text{u}} = 10$ dB, the spectral efficiency and the energy efficiency reduce by factors of 3 and 2.7, respectively. However, with low transmit power where the spectral efficiency is smaller than 10 bits/s/Hz, the system performance is not affected much by the pilot contamination. Furthermore, we can see that in a multicell scenario with high pilot contamination, MRC achieves a better performance than ZF.

6 Conclusion

Very large MIMO systems offer the opportunity of increasing the spectral efficiency (in terms of bits/s/Hz sum-rate in a given cell) by one or two orders of magnitude, and simultaneously improving the energy efficiency (in terms of bits/J) by three orders of magnitude. This is possible with simple linear processing such as MRC or ZF at the BS, and using channel estimates obtained from uplink pilots even in a high mobility environment where half of the channel coherence interval is used for training. Generally, ZF outperforms MRC owing to its ability to cancel intracell interference. However, in multicell environments with strong pilot contamination,

this advantage tends to diminish. MRC has the additional benefit of facilitating a distributed per-antenna implementation of the detector. Quantitatively, with MRC, 100 antennas can serve about 50 terminals in the same time-frequency resource, each terminal having a fading-free throughput of about 1 bpcu, and hence the system offering a sum-throughput of about 50 bpcu. These conclusions are valid under a channel model that includes the effects of small-scale Rayleigh fading, but neglects the effects of large-scale fading (see the discussion after (64)).

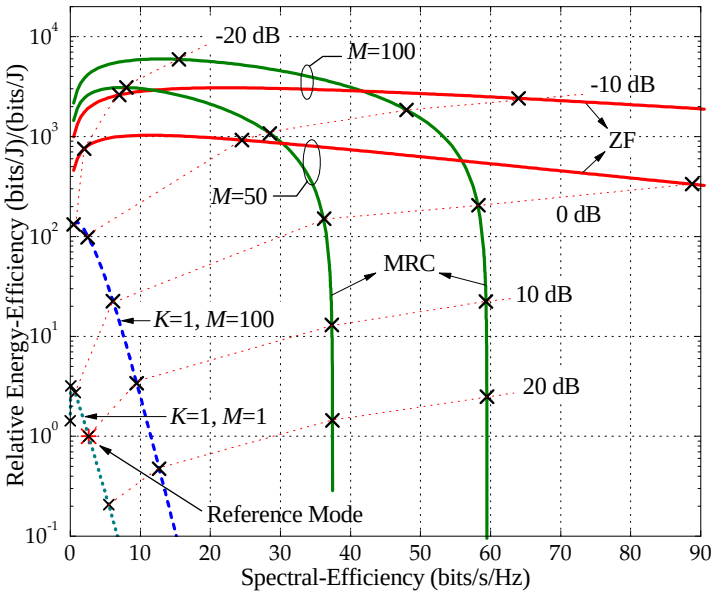


Figure 6: Energy efficiency (normalized with respect to the reference mode) versus spectral efficiency for MRC and ZF receiver processing with imperfect CSI. The reference mode corresponds to $K = 1, M = 1$ (single antenna, single user), and a transmit power of $p_{\mathrm{u}} = 10$ dB. The coherence interval is $T = 196$ symbols. For the dashed curves (marked with $K = 1$), the transmit power p_{u} and the fraction of the coherence interval τ/T spent on training was optimized in order to maximize the energy efficiency for a fixed spectral efficiency. For the green and red curves (marked MRC and ZF; shown for $M = 50$ and $M = 100$ antennas, respectively), the number of users K was optimized jointly with p_{u} and τ/T to maximize the energy efficiency for given spectral efficiency. Any operating point on the curves can be obtained by appropriately selecting p_{u} and optimizing with respect to K and τ/T. The number marked next to the $\times$ marks on each curve is the power p_{u} spent by the transmitter.

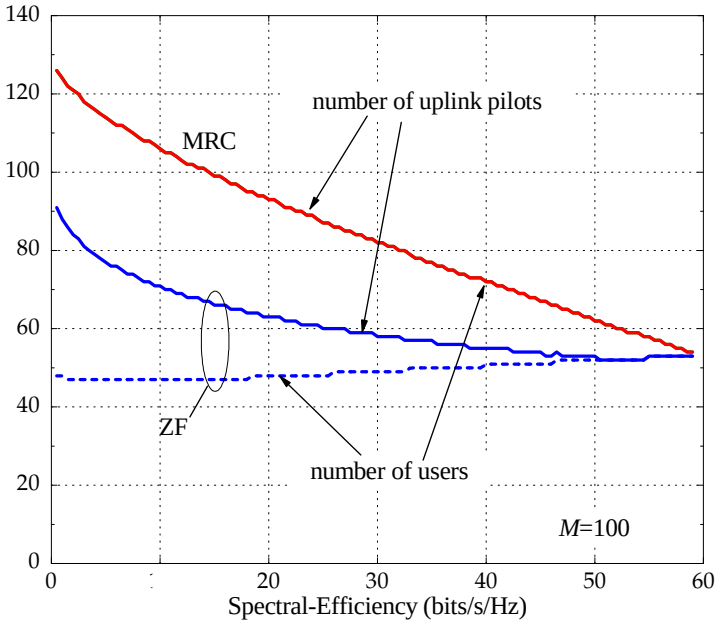


Figure 7: Optimal number of users K and number of symbols τ spent on training, out of a total of $T = 196$ symbols per coherence interval, for the curves in Fig. 5 corresponding to $M = 100$ antennas.

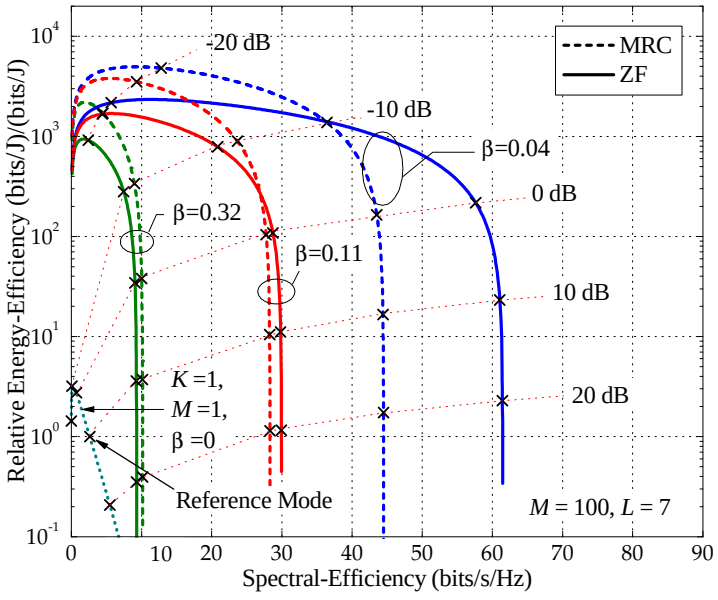


Figure 8: Same as Figure 5, but for a multicell scenario, with $L = 7$ cells, and coherence interval $T = 196$.

Appendix

A Proof of Proposition 2

From (16), we have

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mrc}} = \log_2\left(1+\left(\mathbb{E}\left\{\frac{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\tilde{g}_i|^2+1}{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^2}\right\}\right)^{-1}\right), \tag{76}$$

where $\tilde{g}_i \triangleq \frac{\boldsymbol{g}_k^H\boldsymbol{g}_i}{\|\boldsymbol{g}_k\|}$. Conditioned on $\boldsymbol{g}_k$, $\tilde{g}_i$ is a Gaussian RV with zero mean and variance β_i which does not depend on $\boldsymbol{g}_k$. Therefore, $\tilde{g}_i$ is Gaussian distributed and independent of $\boldsymbol{g}_k$, $\tilde{g}_i \sim \mathcal{CN}(0,\beta_i)$. Then,

$$\begin{aligned}\mathbb{E}\left\{\frac{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\tilde{g}_i|^2+1}{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^2}\right\} &= \left(p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}\mathbb{E}\left\{|\tilde{g}_i|^2\right\}+1\right)\mathbb{E}\left\{\frac{1}{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^2}\right\}\\ &= \left(p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}\beta_i+1\right)\mathbb{E}\left\{\frac{1}{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^2}\right\}. \end{aligned}\tag{77}$$

Using the identity [20]

$$\mathbb{E}\left\{\mathtt{tr}\left(\boldsymbol{W}^{-1}\right)\right\} = m/(n-m), \tag{78}$$

where $\boldsymbol{W}\sim\mathcal{W}_m(n,\boldsymbol{I}_n)$ is an $m\times m$ central complex Wishart matrix with n $(n>m)$ degrees of freedom, we obtain

$$\mathbb{E}\left\{\frac{1}{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^2}\right\} = \frac{1}{p_{\mathrm{u}}(M-1)\beta_k},\ \text{for } M\geq 2. \tag{79}$$

Substituting (79) into (77), we arrive at the desired result (17).

B Proof of Proposition 3

From (3), we have

$$\begin{aligned}\mathbb{E}\left\{\left[\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}\right\} &= \frac{1}{\beta_k}\,\mathbb{E}\left\{\left[\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\right]_{kk}\right\} \\ &= \frac{1}{K\beta_k}\,\mathbb{E}\left\{\mathtt{tr}\left[\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\right]\right\} \\ &\overset{(a)}{=} \frac{1}{(M-K)\,\beta_k},\ \text{for } M \geq K+1, \end{aligned} \tag{80}$$

where (a) is obtained by using (78). Using (80), we get (21).

References

[1] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Sig. Proc. Mag.*, vol. 24, no. 5, pp. 36–46, 2007.

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

[3] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[4] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.

[5] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality" *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

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

[7] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of BS antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[8] ——, "How much training is required for multiuser MIMO," in *Fortieth Asilomar Conference on Signals, Systems and Computers (ACSSC '06)*, Pacific Grove, CA, USA, Oct. 2006, pp. 359–363.

[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Proc. Mag.*, accepted. [Online]. Available: arxiv.org/abs/1201.3210.

[10] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?," in *Proc. 49th Allerton Conference on Communication, Control, and Computing*, 2011.

[11] A. Fehske, G. Fettweis, J. Malmodin and G. Biczok, "The global footprint of mobile communications: the ecological and economic perspective," *IEEE Communications Magazine*, pp. 55-62, August 2011.

[12] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.

[13] H. Huh, G. Caire, H. C. Papadopoulos, S. A. Rampshad, "Achieving large spectral efficiency with TDD and not-so-many base station antennas," in *Proc. IEEE Antennas and Propagation in Wireless Communications (APWC)*, 2011.

[14] 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," in *Proc. IEEE Int. Works. Signal Process. Adv. Wireless Commun. (SPAWC)*, 2010.

[15] H. Cramér, *Random Variables and Probability Distributions*. Cambridge, UK: Cambridge University Press, 1970.

[16] N. Kim and H. Park, "Performance analysis of MIMO system with linear MMSE receiver," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4474–4478, Nov. 2008.

[17] H. Gao, P. J. Smith, and M. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666–672, May 1998.

[18] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 271–286, Jan. 2006.

[19] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA: Academic, 2007.

[20] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

[21] A. Lozano, R. W. Heath Jr., and J. G. Andrews, "Fundamentral limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226 , Mar. 2013.

Paper B

The Multicell Multiuser MIMO Uplink with Very Large Antenna Arrays and a Finite-Dimensional Channel

Refereed article published in the IEEE Transactions on
Communications 2013.

The Multicell Multiuser MIMO Uplink with Very Large Antenna Arrays and a Finite-Dimensional Channel

Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta

Abstract

We consider multicell multiuser MIMO systems with a very large number of antennas at the base station (BS). We assume that the channel is estimated by using uplink training. We further consider a physical channel model where the angular domain is separated into a finite number of distinct directions. We analyze the so-called pilot contamination effect discovered in previous work, and show that this effect persists under the finite-dimensional channel model that we consider. In particular, we consider a uniform array at the BS. For this scenario, we show that when the number of BS antennas goes to infinity, the system performance under a finite-dimensional channel model with P angular bins is the same as the performance under an uncorrelated channel model with P antennas. We further derive a lower bound on the achievable rate of uplink data transmission with a linear detector at the BS. We then specialize this lower bound to the cases of maximum-ratio combining (MRC) and zero-forcing (ZF) receivers, for a finite and an infinite number of BS antennas. Numerical results corroborate our analysis and show a comparison between the performances of MRC and ZF in terms of sum-rate.

1 Introduction

Multiuser multiple-input multiple-output (MU-MIMO) systems, where several co-channel users communicate with a base station (BS) equipped with multiple antennas, have recently attracted substantial interest. Such systems can offer a spatial multiplexing gain even if the users have only a single antenna each [1–3]. Most studies have assumed that the BS has some channel state information (CSI). The problem of not having a priori CSI at the BS was considered in [4–6], assuming that the channel estimation is done by using uplink pilots. References [4,5] considered a single-cell setting. This is only reasonable when the pilot sequences used in each cell are orthogonal to those used in other cells. However, in practical cellular networks, the channel coherence time typically is not long enough to allow for orthogonality between the pilots in different cells. Therefore, non-orthogonal training sequences must be utilized and hence, the multicell setting should be considered. In the multicell scenario with non-orthogonal pilots in different cells, channel estimates obtained in a given cell will be impaired by pilots transmitted by users in other cells. This effect, called "pilot contamination", has been analyzed in [7].

Conventional MIMO technology uses a relatively small number of antennas at the BS. The LTE standard, for example, allows for up to 8 antenna ports. In this paper, by contrast, we are concerned with MIMO systems that use a very large number of antennas at the BS compared to systems being built today, i.e., a hundred or more antennas [8,9]. With a very large antenna array, the transmit power can be reduced by an order of magnitude, or more. For example, to obtain the same quality-of-service as with a single-antenna BS, a 100-antenna array would need to transmit with only 1% of the power [10]. A fundamental consequence of the number of antennas growing large is that things that were random before become deterministic. In particular, the effect of thermal noise and small scale fading is averaged out. In [8], the author considered multicell MU-MIMO systems with very large antenna arrays at the BS and showed that with simple maximum-ratio combining (MRC) for the uplink, and maximum-ratio transmission for the downlink, when the number of antennas increases without bound, uncorrelated noise, fast fading, and intracell interference vanish. Instead, the pilot contamination effect discussed above dictates the ultimate limit on the system performance. To illustrate with a quantitative result, the author in [8] showed that for an unlimited number of BS antennas, in a multicell MU-MIMO with a frequency reuse factor of 7, and a bandwidth of 20 MHz, each user can achieve a downlink link average net throughput of 17 Mbits/sec.

The complexity issue in practical systems is a growing concern. For very large MIMO systems, all the complexity is at the BS. The vision is that the large antenna array can be built from very simple and inexpensive antenna units. The signal processing should be simple, e.g., using linear precoders and linear detectors. Among linear detectors, MRC has an advantage since it can be implemented in a distributed manner, i.e., each antenna performs multiplication of the received signals with the conjugate transpose of the channel, without sending the entire baseband

signal to the BS for processing. Very recently, the design and analysis of such very large MIMO systems has regained significant interest [8–14]. Important practical aspects of using very many antennas have been discovered. For example, in [10], the authors showed that with simple linear detectors at the BS for the uplink, by using a very large antenna array, the transmit power of each user can be made inversely propositional to the number of BS antennas for perfect CSI, and to the square-root of the number of BS antennas for imperfect CSI with no performance degradation. Note that operating with $M \gg K$, where M is the number of BS antennas and K is the number of users, is both a desirable and a natural operating point since K is limited by mobility (approximately the coherence interval divided by the channel delay-spread). But M can be made as big as desired, and with TDD CSI acquisition there is no overhead penalty with respect to M. Taken together, further studies of very large MIMO systems and implementation aspects of such systems are well motivated.

Most of the studies referred to above assume that the channels are independent [4, 5, 7] or that the channel vectors for different users are asymptotically orthogonal [8, 10]. However, in reality, the channel vectors for different users are generally correlated, or not asymptotically orthogonal, and can be modelled as P-dimensional, where P is the number of angular bins. This is so because the antennas are not sufficiently well separated or the propagation environment does not offer rich enough scattering. In many scenarios, P is large. However, in some specific scenarios, P is small, for example for keyhole channels [15]. In our work, we investigate the performance of large antenna systems in the regime where P is much smaller than the number of BS antennas. When $M \gg P$, then the system is saturated with respect to throughput gains, but there will still be radiated energy-efficiency gains for arbitrarily large M (at least until the array gets so big that it begins to envelope the P scatters).

1.1 Contributions

We investigate the performance of multicell MU-MIMO with large antenna arrays under a physical channel model. More precisely, we consider a channel model in which the angular domain is partitioned into a large, but finite number of directions which is smaller than the number of BS antennas. The channels are estimated by using uplink training. For such channels, the number of parameters to be estimated is fixed regardless of the number of antennas. The paper makes the following specific contributions:

- We show that the pilot contamination effect persists under a finite-dimensional channel model. When the number of BS antennas grows without bound and the antenna array is uniform, the system performance under a finite-dimensional model with P angular bins is the same as the performance under an uncorrelated channel model with P antennas.

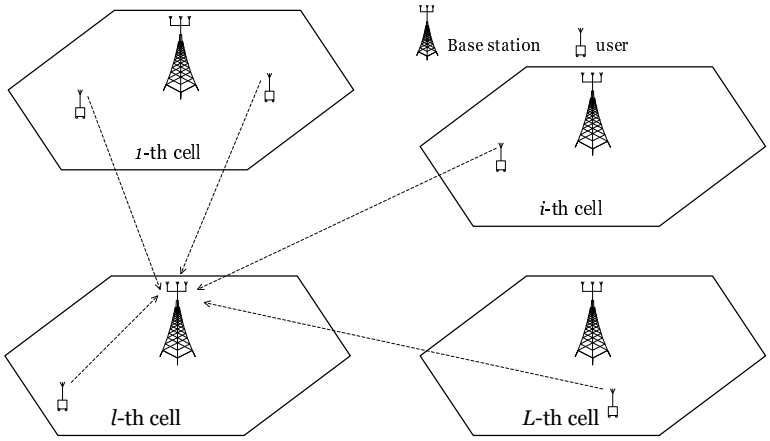


Figure 1: Uplink transmission in multi-cell multi-user MIMO systems. The lth BS receives signals from all users in all cells.

- We derive novel closed-form lower bounds on the achievable rates for the uplink transmission, assuming MRC and zero-forcing (ZF) processing at the BS. These bounds are valid for a large, but finite, number of antennas. We then compare the performances of MRC and ZF for different propagation parameters, reuse factors, path loss exponents, and cell radii.

1.2 Notation

The superscripts T, $*$, and $\dagger$ stand for the transpose, conjugate, and conjugate-transpose, respectively. $[\boldsymbol{A}]_{ij}$ represents the (i,j)th entry of a matrix $\boldsymbol{A}$. Finally, we use $\tilde{\mathcal{N}}_{m,n}(\boldsymbol{M},\boldsymbol{\Sigma},\boldsymbol{\Psi})$ to denote a matrix-variate complex Gaussian distribution with mean matrix $\boldsymbol{M} \in \mathbb{C}^{m\times n}$ and covariance matrix $\boldsymbol{\Psi}^T \otimes \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} \in \mathbb{C}^{m\times m}$, $\boldsymbol{\Psi} \in \mathbb{C}^{n\times n}$, and $\otimes$ denotes the Kronecker product.

2 System Model

2.1 Multi-cell Multi-user MIMO Model

Consider L cells, where each cell contains one BS equipped with M antennas and K single-antenna users. Assume that the L BSs share the same frequency band.

We consider uplink transmission, where the lth BS receives signals from all users in all cells. See Fig. 1. Then, the $M \times 1$ received vector at the lth BS is given by

$$\boldsymbol{y}_l = \sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \boldsymbol{\Upsilon}_{il} \boldsymbol{x}_i + \boldsymbol{n}_l, \tag{1}$$

where $\boldsymbol{\Upsilon}_{il}$ represents the $M \times K$ channel matrix between the lth BS and the K users in the ith cell, i.e., $[\boldsymbol{\Upsilon}_{il}]_{mk}$ is the channel coefficient between the mth antenna of the lth BS and the kth user in the ith cell; $\sqrt{p_{\mathrm{u}}}\boldsymbol{x}_i$ is the $K \times 1$ transmitted vector the K users in the ith cell (the average power used by each user is p_{u}); and $\boldsymbol{n}_l$ contains additive white Gaussian noise (AWGN). We assume that the elements of $\boldsymbol{n}_l$ are Gaussian with zero mean and unit variance.

2.2 Physical Channel Model

The performance of MIMO systems depends critically on the complexity of the propagation environment and the properties of the antenna arrays being used, and ultimately on the number of degrees of freedom offered by the physical channels. In practice, the dimension of the physical channel is finite [16, 17]. Therefore, here we introduce a finite-dimensional channel model, which will be used throughout the paper. In this model, the angular domain is divided into a large but finite number of directions P, which is fixed regardless of the number of BS antennas. Our analysis will require $P \leq M$. Each direction, corresponding to the angle ϕ_p, $\phi_p \in [-\pi/2, \pi/2]$, $p = 1, ..., P$, is associated with an $M \times 1$ array steering vector $\boldsymbol{a}(\phi_p)$:

$$\boldsymbol{a}(\phi_p) = \frac{1}{\sqrt{P}} \left[e^{-jf_1(\phi_p)}, e^{-jf_2(\phi_p)}, ..., e^{-jf_M(\phi_p)} \right]^T, \tag{2}$$

where $f_m(\phi)$ is a function of ϕ. The channel vector from the kth user in the ith cell to the lth BS is then a linear combination of P steering vectors as follows: $\sum_{p=1}^{P} g_{ilkp}\boldsymbol{a}(\phi_p)$, where g_{ilkp} is the propagation coefficient from the kth user in the ith cell to the lth BS, associated with the pth physical direction (with direction-of-arrival ϕ_p). The factor $\frac{1}{\sqrt{P}}$ in (2) is used to normalize the channel. Let $\boldsymbol{G}_{il} \triangleq [\boldsymbol{g}_{il1} \cdots \boldsymbol{g}_{ilK}]$ be a $P \times K$ matrix with $\boldsymbol{g}_{ilk} \triangleq [g_{ilk1} \cdots g_{ilkP}]^T$ that contains the propagation coefficients from the kth user in the ith cell to the lth BS. Then, the channel matrix between the lth BS and K users in the ith cell is

$$\boldsymbol{\Upsilon}_{il} = \boldsymbol{A}\boldsymbol{G}_{il}, \tag{3}$$

where $\boldsymbol{A} \triangleq [\boldsymbol{a}(\phi_1) \cdots \boldsymbol{a}(\phi_P)]$ is a full rank $M \times P$ matrix. We stress that $\boldsymbol{A}$ is fixed and known whereas $\boldsymbol{G}_{il}$ has to be estimated at the BS.

The propagation channel $\boldsymbol{G}_{il}$ models independent fast fading, geometric attenuation, and log-normal shadow fading. Its elements g_{ilkp} are assumed to be independent, and given by

$$g_{ilkp} = h_{ilkp}\sqrt{\beta_{ilk}}, \quad p = 1, 2, ..., P, \tag{4}$$

where h_{ilkp} is the fast fading coefficient from the kth user in the ith cell to the lth BS, associated with the pth physical direction. The coefficient h_{ilkp} is assumed to be zero-mean and have unit variance. Moreover, $\sqrt{\beta_{ilk}}$ models the geometric attenuation and shadow fading which are assumed to be independent of the direction p and to be constant and known a priori. This assumption is reasonable since the value of β_{ilk} changes very slowly with time. Then, we have

$$\boldsymbol{G}_{il} = \boldsymbol{H}_{il}\boldsymbol{D}_{il}^{1/2}, \tag{5}$$

where $\boldsymbol{H}_{il}$ is the $P \times K$ matrix of fast fading coefficients between the K users in the ith cell and the lth BS, i.e., $[\boldsymbol{H}_{il}]_{kp} = h_{ilkp}$, and $\boldsymbol{D}_{il}$ is a $K \times K$ diagonal matrix whose diagonal elements are given by $[\boldsymbol{D}_{il}]_{kk} = \beta_{ilk}$. Therefore, (3) can be written as

$$\boldsymbol{y}_l = \sqrt{p_{\mathrm{u}}}\boldsymbol{A}\sum_{i=1}^{L}\boldsymbol{G}_{il}\boldsymbol{x}_i + \boldsymbol{n}_l = \sqrt{p_{\mathrm{u}}}\boldsymbol{A}\sum_{i=1}^{L}\boldsymbol{H}_{il}\boldsymbol{D}_{il}^{1/2}\boldsymbol{x}_i + \boldsymbol{n}_l. \tag{6}$$

3 Channel Estimation

Channel estimation is performed by using training sequences received on the uplink. We assume that the BS uses minimum mean-square-error (MMSE) estimation.

3.1 Uplink Training

We assume that an interval of length τ symbols is used for uplink training, and that this interval is shorter than the coherence time of the channel. All users in all cells simultaneously transmit pilot sequences of length τ symbols. The assumption on synchronized transmission represents the worst case from the pilot contamination point of view. The reason for this was explained in detail in [8, Section VII-G]. Suppose that the users in the lth cell are transmitting uplink pilots at the same time that the users in other cells are transmitting data. Whatever data is transmitted in other cells, it can always be expanded in terms of the orthogonal pilot sequences that are transmitted in the lth cell, so pilot contamination ensues. The point is that, whatever the users in other cells transmit, it appears as a coherent signal across the M antennas in the lth cell.

We further assume that the same set of pilot sequences is used in all L cells. This assumption makes no fundamental difference compared with using different pilot sequences in different cells [8, Section VII-F]. The pilot sequences used in the lth cell can be represented by a $\tau \times K$ matrix $\sqrt{p_\mathrm{p}}\boldsymbol{\Phi}_l = \sqrt{p_\mathrm{p}}\boldsymbol{\Phi}$ ($\tau \geq K$), which satisfies $\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi} = \boldsymbol{I}_K$, where $p_\mathrm{p} = \tau p_\mathrm{u}$. From (6), the received pilot matrix at the lth BS is

$$\boldsymbol{Y}_{\mathrm{p},l} = \sqrt{p_\mathrm{p}}\boldsymbol{A}\sum_{i=1}^{L}\boldsymbol{G}_{il}\boldsymbol{\Phi}^T + \boldsymbol{N}_l = \sqrt{p_\mathrm{p}}\boldsymbol{A}\sum_{i=1}^{L}\boldsymbol{H}_{il}\boldsymbol{D}_{il}^{1/2}\boldsymbol{\Phi}^T + \boldsymbol{N}_l, \tag{7}$$

where $\boldsymbol{N}_l$ is an $M \times \tau$ complex AWGN matrix, $\boldsymbol{N}_l \sim \tilde{\mathcal{N}}_{M,\tau}(\boldsymbol{0}, \boldsymbol{I}_M, \boldsymbol{I}_\tau)$.

3.2 Minimum Mean-Square Error Channel Estimation

Since the projection of the received pilot $\boldsymbol{Y}_{\mathrm{p},l}$ on $\boldsymbol{\Phi}^*$ is a sufficient statistic for the estimation of $\boldsymbol{H}_{ll}$ [18], we use $\tilde{\boldsymbol{Y}}_{\mathrm{p},l} \triangleq \boldsymbol{Y}_{\mathrm{p},l}\boldsymbol{\Phi}^*$ to estimate the channel. We have

$$\tilde{\boldsymbol{Y}}_{\mathrm{p},l} = \sqrt{p_\mathrm{p}}\boldsymbol{A}\sum_{i=1}^{L}\boldsymbol{H}_{il}\boldsymbol{D}_{il}^{1/2} + \boldsymbol{W}_l, \tag{8}$$

where $\boldsymbol{W}_l \triangleq \boldsymbol{N}_l\boldsymbol{\Phi}^*$, $\boldsymbol{W}_l \sim \tilde{\mathcal{N}}_{M,K}(\boldsymbol{0}, \boldsymbol{I}_M, \boldsymbol{I}_K)$. Since $\boldsymbol{H}_{ll}$ has independent columns, we can estimate each column of $\boldsymbol{H}_{ll}$ independently. Let $\tilde{\boldsymbol{y}}_{\mathrm{p},lk}$ be the kth column of $\tilde{\boldsymbol{Y}}_{\mathrm{p},l}$. Then

$$\tilde{\boldsymbol{y}}_{\mathrm{p},lk} = \sqrt{p_\mathrm{p}}\boldsymbol{A}\boldsymbol{h}_{llk}\sqrt{\beta_{llk}} + \sqrt{p_\mathrm{p}}\boldsymbol{A}\sum_{i\neq l}^{L}\boldsymbol{h}_{ilk}\sqrt{\beta_{ilk}} + \boldsymbol{w}_{lk}, \tag{9}$$

where $\boldsymbol{h}_{ilk}$ and $\boldsymbol{w}_{lk}$ are the kth columns of $\boldsymbol{H}_{il}$ and $\boldsymbol{W}_l$, respectively. From (9), we can see that the interference-plus-noise term, $\sqrt{p_\mathrm{p}}\boldsymbol{A}\sum_{i\neq l}^{L}\boldsymbol{h}_{ilk}\sqrt{\beta_{ilk}} + \boldsymbol{w}_{lk}$, is Gaussian distributed. Therefore, using MMSE estimation is optimal. The MMSE estimate of $\boldsymbol{h}_{llk}$ is given by [19]

$$\hat{\boldsymbol{h}}_{llk} = \sqrt{p_\mathrm{p}\beta_{llk}}\left(p_\mathrm{p}\boldsymbol{A}^\dagger\boldsymbol{A}\sum_{i=1}^{L}\beta_{ilk} + \boldsymbol{I}_P\right)^{-1}\boldsymbol{A}^\dagger\tilde{\boldsymbol{y}}_{\mathrm{p},lk}. \tag{10}$$

Since $\boldsymbol{A}$ is a matrix whose pth column is given by (2), the pth diagonal element of $p_\mathrm{p}\boldsymbol{A}^\dagger\boldsymbol{A}\sum_{i=1}^{L}\beta_{ilk}$ in (10) equals $\frac{Mp_\mathrm{p}}{P}\sum_{i=1}^{L}\beta_{ilk}$. The uplink is typically interference-limited, so $\frac{Mp_\mathrm{p}}{P}\sum_{i=1}^{L}\beta_{ilk} \gg 1$. Therefore, $\hat{\boldsymbol{h}}_{llk}$ can be approximated as

$$\hat{\boldsymbol{h}}_{llk} \approx \sqrt{p_\mathrm{p}\beta_{llk}}\left(p_\mathrm{p}\boldsymbol{A}^\dagger\boldsymbol{A}\sum_{i=1}^{L}\beta_{ilk}\right)^{-1}\boldsymbol{A}^\dagger\tilde{\boldsymbol{y}}_{\mathrm{p},lk}. \tag{11}$$

Thus, the MMSE estimate of $\boldsymbol{H}_{ll}$ is

$$\hat{\boldsymbol{H}}_{ll} = \frac{1}{\sqrt{p_{\mathrm{p}}}} \left(\boldsymbol{A}^{\dagger}\boldsymbol{A}\right)^{-1} \boldsymbol{A}^{\dagger}\tilde{\boldsymbol{Y}}_{\mathrm{p},l}\boldsymbol{D}_{l}^{-1}\boldsymbol{D}_{ll}^{1/2}, \tag{12}$$

where $\boldsymbol{D}_l \triangleq \sum_{i=1}^{L} \boldsymbol{D}_{il}$ is a $K \times K$ diagonal matrix whose kth diagonal element equals $\sum_{i=1}^{L} \beta_{ilk}$. Then, the estimate of the physical channel matrix between the lth BS and the K users in the lth cell is given by

$$\hat{\boldsymbol{\Upsilon}}_{ll} = \boldsymbol{A}\hat{\boldsymbol{H}}_{ll}\boldsymbol{D}_{ll}^{1/2} = \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{\Pi}_{\boldsymbol{A}}\tilde{\boldsymbol{Y}}_{\mathrm{p},l}\boldsymbol{D}_{l}^{-1}\boldsymbol{D}_{ll}, \tag{13}$$

where $\boldsymbol{\Pi}_{\boldsymbol{A}} \triangleq \boldsymbol{A}\left(\boldsymbol{A}^{\dagger}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\dagger}$ is the orthogonal projection onto $\boldsymbol{A}$. We can see that since post-multiplication of $\boldsymbol{Y}_{\mathrm{p},l}$ with $\boldsymbol{\Phi}^{*}$ means just multiplication with the pseudoinverse ($\boldsymbol{\Phi}^{\dagger}\boldsymbol{\Phi} = \boldsymbol{I}_K$). Note that $\tilde{\boldsymbol{Y}}_{\mathrm{p},l}$ in (8) is the conventional least-squares channel estimate. The optimal channel estimator that we derived thus performs conventional channel estimation and then projects the estimate onto the physical (beamspace) model for the array. Substituting (8) into (13), we obtain

$$\hat{\boldsymbol{\Upsilon}}_{ll} = \boldsymbol{A}\sum_{i=1}^{L}\boldsymbol{H}_{il}\boldsymbol{D}_{il}^{1/2}\boldsymbol{D}_{l}^{-1}\boldsymbol{D}_{ll} + \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{\Pi}_{\boldsymbol{A}}\boldsymbol{W}_{l}\boldsymbol{D}_{l}^{-1}\boldsymbol{D}_{ll}. \tag{14}$$

Note that, the number of parameters to estimate (elements of $\boldsymbol{H}_{ll}$) is fixed regardless M, and the number of observations (elements of $\boldsymbol{Y}_{\mathrm{p},l}$) increases with M, and goes to infinity when $M \to \infty$. However, owing to the finite dimensionality of the channel, the number of linearly independent observations is also finite. In particular, we can see from (8) that the effective number of observations is $P \times \tau$ which is fixed regardless of M. Therefore, we cannot estimate the channel arbitrary accurately by increasing the number of antennas.

We can see that the channel estimate of the lth BS (for the K users in the lth cell) includes contributions from all channel vectors from other cells to the lth BS. This causes the pilot contamination. Note that the pilot contamination effect is fundamental and does not result as a deficiency of the procedure used for channel estimation. The received signal at the BS is a linear combination of all transmitted signals from all users in all cells. The physical properties of the desired and interference signals are the same, and we cannot distinguish them. Therefore, the pilot contamination will exist regardless of which channel estimation technique that is used.

4 Analysis of Uplink Data Transmission

In this section, we analyze the achievable rates on the uplink for a finite and an infinite number of BS antennas. We consider the model in (3). From (3), we

can see that when considering large-scale MIMO systems, the interference term, $\sqrt{p_{\mathrm{u}}}\sum_{i\neq l}^{L}\boldsymbol{\Upsilon}_{il}\boldsymbol{x}_i$, can be approximated as Gaussian distributed by using the Cramér Central Limit Theorem (CLT) [20]. Hence, conditioned on $\boldsymbol{\Upsilon}_{ll}$ and $\boldsymbol{x}_l$, $\boldsymbol{y}_l$ is approximately Gaussian distributed with mean $\sqrt{p_{\mathrm{u}}}\boldsymbol{\Upsilon}_{ll}\boldsymbol{x}_l$ and covariance $\boldsymbol{R}_l$, where

$$\boldsymbol{R}_l = \mathtt{Cov}\left\{\sqrt{p_{\mathrm{u}}}\sum_{i\neq l}^{L}\boldsymbol{\Upsilon}_{il}\boldsymbol{x}_i + \boldsymbol{n}_l\right\} = p_{\mathrm{u}}\sum_{i\neq l}^{L}\sum_{k=1}^{K}\beta_{ilk}\boldsymbol{A}\boldsymbol{A}^{\dagger} + \boldsymbol{I}_M. \tag{15}$$

Therefore, the "mismatched" detector that treats the estimated channel as the true one is given by

$$\hat{\boldsymbol{x}}_l = \arg\min_{\boldsymbol{x}_l\in\mathcal{X}}\left(\boldsymbol{y}_l - \sqrt{p_{\mathrm{u}}}\hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{x}_l\right)^{\dagger}\boldsymbol{R}_l^{-1}\left(\boldsymbol{y}_l - \sqrt{p_{\mathrm{u}}}\hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{x}_l\right). \tag{16}$$

We next derive the optimal detector for the lth BS that takes the channel estimation errors into account when performing data detection. Let $\tilde{\boldsymbol{\Upsilon}}_{ll} \triangleq \hat{\boldsymbol{\Upsilon}}_{ll} - \boldsymbol{\Upsilon}_{ll}$ be the channel estimation error. From the properties of MMSE estimation, $\tilde{\boldsymbol{\Upsilon}}_{ll}$ is independent of $\hat{\boldsymbol{\Upsilon}}_{ll}$. By the CLT, for large-scale systems, conditioned on $\hat{\boldsymbol{\Upsilon}}_{ll}$ and $\boldsymbol{x}_l$, $\boldsymbol{y}_l$ is approximately Gaussian distributed with mean $\sqrt{p_{\mathrm{u}}}\hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{x}_l$ and covariance $\hat{\boldsymbol{R}}_l$, where

$$\begin{aligned}\hat{\boldsymbol{R}}_l &= \mathtt{Cov}\left\{-\sqrt{p_{\mathrm{u}}}\tilde{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{x}_l + \sqrt{p_{\mathrm{u}}}\sum_{i\neq l}^{L}\boldsymbol{\Upsilon}_{il}\boldsymbol{x}_i + \boldsymbol{n}_l\right\}\\ &\approx p_{\mathrm{u}}\left(\boldsymbol{x}_l^T\boldsymbol{D}\boldsymbol{x}_l^*\right)\boldsymbol{A}\boldsymbol{A}^{\dagger} + p_{\mathrm{u}}\sum_{i\neq l}^{L}\sum_{k=1}^{K}\beta_{ilk}\boldsymbol{A}\boldsymbol{A}^{\dagger} + \boldsymbol{I}_M,\end{aligned} \tag{17}$$

where $\boldsymbol{D} \triangleq \mathrm{diag}\left(d_1, d_2, ..., d_K\right)$, $d_k \triangleq \frac{\beta_{llk}\sum_{i\neq l}^{L}\beta_{ilk}}{\sum_{i=1}^{L}\beta_{ilk}}$, and the approximation is obtained by using (11). Under the assumption that the transmitted signal has constant modulus so that $|x_{lk}|^2 = 1$ (i.e. x_{lk} comes from an M-PSK constellation), we have

$$\hat{\boldsymbol{R}}_l = p_{\mathrm{u}}\sum_{k=1}^{K}\left(d_k + \sum_{i\neq l}^{L}\beta_{ilk}\right)\boldsymbol{A}\boldsymbol{A}^{\dagger} + \boldsymbol{I}_M. \tag{18}$$

Since $\hat{\boldsymbol{R}}_l$ does not depend on $\boldsymbol{x}_l$, the maximum-likelihood detector is

$$\hat{\boldsymbol{x}}_l = \arg\min_{\boldsymbol{x}_l\in\mathcal{X}}\left(\boldsymbol{y}_l - \sqrt{p_{\mathrm{u}}}\hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{x}_l\right)^{\dagger}\hat{\boldsymbol{R}}_l^{-1}\left(\boldsymbol{y}_l - \sqrt{p_{\mathrm{u}}}\hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{x}_l\right). \tag{19}$$

From (16), (19), and using the fact that $\boldsymbol{R}_l^{-1} \approx \hat{\boldsymbol{R}}_l^{-1}$ for large MU-MIMO systems, we can see that, with imperfect CSI, the gap between the performances of the detector which takes the channel estimation errors into account and the one which treats the channel estimate as the true channel is very small. Therefore, in our

analysis, we assume that the BS treats the channel estimate obtained by uplink training as the true channel. It uses this channel estimate to detect the signals transmitted by the K users in its cell.

Theoretically, the maximum-likelihood multiuser detector can be used to obtain optimal performance. However, this scheme has a complexity which is exponential in the number of users. Therefore, we consider linear detection schemes at the BS, which reduce the decoding complexity by separating the transmitted signal streams, and then decoding each stream independently. Let $\boldsymbol{F}_{ll}$ be an $M \times K$ linear detection matrix which depends on the channel estimate $\hat{\boldsymbol{\Upsilon}}_{ll}$. The lth BS processes its received signal by multiplying it by $\boldsymbol{F}_{ll}^{\dagger}$ as follows

$$\boldsymbol{r}_l = \boldsymbol{F}_{ll}^{\dagger}\boldsymbol{y}_l = \boldsymbol{F}_{ll}^{\dagger}\left(\sqrt{p_{\mathrm{u}}}\sum_{i=1}^{L}\boldsymbol{\Upsilon}_{il}\boldsymbol{x}_i + \boldsymbol{n}_l\right). \tag{20}$$

Linear detectors are known to perform exceedingly well in large MU-MIMO systems [10].

In [8], the author showed that as M grows without bound and under the assumption of asymptotically orthogonal channel vectors, even with simple processing such as MRC, the effects of uncorrelated noise and small scale fading vanish. The only remaining impairment stems from the pilot contamination. Indeed, this effect constitutes an ultimate limit on the performance of multicell MU-MIMO systems. This raises the question of how fundamentally does a finite-dimensional channel model change the nature and the performance of such a system. In particular, does the pilot contamination effect persist under a finite-dimensional channel model? To answer this, we first consider the effect of pilot contamination for two conventional linear detection schemes, MRC and ZF receivers, for an infinite number of BS antennas with a finite-dimensional channel model. Then, to gain further insight into the pilot contamination effect, we derive lower bounds on the achievable rate for an infinite number of BS antennas of these linear detectors. We also derive a lower bound on the achievable rate for a finite but large M.

4.1 The Pilot Contamination Effect

4.1.1 MRC Receiver

The MRC technique linearly combines the data transmitted from all users in order to maximize the signal-to-noise ratio (SNR). For this technique, the linear detection matrix is the channel estimate, i.e., $\boldsymbol{F}_{ll} = \hat{\boldsymbol{\Upsilon}}_{ll}$. From (17), (20), we have

$$\boldsymbol{r}_l = \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{D}_{ll}\boldsymbol{D}_l^{-1}\left(\sqrt{p_{\mathrm{p}}}\boldsymbol{A}\sum_{i=1}^{L}\boldsymbol{G}_{il} + \boldsymbol{W}_l\right)^{\dagger}\boldsymbol{\Pi}_{\boldsymbol{A}}\left(\sqrt{p_{\mathrm{u}}}\boldsymbol{A}\sum_{j=1}^{L}\boldsymbol{G}_{jl}\boldsymbol{x}_j + \boldsymbol{n}_l\right). \tag{21}$$

As $M \to \infty$, the products of uncorrelated quantities can be ignored because correlated quantities grow as M while uncorrelated quantities grow only as $\sqrt{M}$ [8]. Then (21) becomes

$$\frac{1}{\sqrt{p_{\mathrm{u}}}M}\boldsymbol{r}_l \to \boldsymbol{D}_{ll}\boldsymbol{D}_l^{-1}\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^{\dagger}\right)\frac{\boldsymbol{A}^{\dagger}\boldsymbol{A}}{M}\left(\sum_{j=1}^{L}\boldsymbol{G}_{jl}\boldsymbol{x}_j\right). \tag{22}$$

We can see that for an unlimited number of antennas, the effect of uncorrelated noise disappears. In particular, the pilot contamination effect, which is due to the interference from users in other cells, persists under the finite-dimensional channel model.

4.1.2 ZF Receiver

The ZF receiver can suppress interuser interference. It has good performance at high SNR. For the ZF technique, we assume that $P \geq K$. The linear detection matrix is the pseudo-inverse matrix of the channel estimate $\hat{\boldsymbol{\Upsilon}}_{ll}^{\dagger}$, i.e., $\boldsymbol{F}_{ll} = \hat{\boldsymbol{\Upsilon}}_{ll}\left(\hat{\boldsymbol{\Upsilon}}_{ll}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{ll}\right)^{-1}$. Similarly to the MRC case, when the number M of BS antennas goes to infinity, only products of correlated quantities remain significant. Then, as M grows large, the received signal after applying the ZF receiver filter becomes

$$\frac{1}{\sqrt{p_{\mathrm{u}}}}\boldsymbol{r}_l \to \boldsymbol{D}_{ll}^{-1}\boldsymbol{D}_l\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^{\dagger}\frac{\boldsymbol{A}^{\dagger}\boldsymbol{A}}{M}\sum_{j=1}^{L}\boldsymbol{G}_{jl} + \frac{\boldsymbol{W}_l^{\dagger}\boldsymbol{\Pi}_{\boldsymbol{A}}\boldsymbol{W}_l}{p_{\mathrm{p}}M}\right)^{-1} \times \left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^{\dagger}\right)\frac{\boldsymbol{A}^{\dagger}\boldsymbol{A}}{M}\left(\sum_{j=1}^{L}\boldsymbol{G}_{jl}\boldsymbol{x}_j\right). \tag{23}$$

Since $M \geq P$ and $\boldsymbol{\Pi}_{\boldsymbol{A}}$ is the orthogonal projection onto $\boldsymbol{A}$, $\boldsymbol{\Pi}_{\boldsymbol{A}}$ can be decomposed as $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\dagger}$, where $\boldsymbol{U}$ is an $M \times M$ unitary matrix, and $\boldsymbol{\Lambda} = \mathrm{diag}\{1, ..., 1, 0, ..., 0\}$ with P diagonal elements equal to 1 and $M - P$ diagonal elements equal to 0. Then

$$\frac{\boldsymbol{W}_l^{\dagger}\boldsymbol{\Pi}_{\boldsymbol{A}}\boldsymbol{W}_l}{p_{\mathrm{p}}M} = \frac{\tilde{\boldsymbol{W}}_l^{\dagger}\boldsymbol{\Lambda}\tilde{\boldsymbol{W}}_l}{p_{\mathrm{p}}M}, \tag{24}$$

where $\tilde{\boldsymbol{W}}_l \triangleq \boldsymbol{U}^{\dagger}\boldsymbol{W}_l \sim \tilde{\mathcal{N}}_{M,K}(\boldsymbol{0}, \boldsymbol{I}_M, \boldsymbol{I}_K)$. Let $\tilde{\boldsymbol{W}}_{lP}$ be the $P \times K$ matrix whose ith row is equal to the ith row of $\tilde{\boldsymbol{W}}_l$. Then $\frac{\boldsymbol{W}_l^{\dagger}\boldsymbol{\Pi}_{\boldsymbol{A}}\boldsymbol{W}_l}{p_{\mathrm{p}}M} = \frac{\tilde{\boldsymbol{W}}_{lP}^{\dagger}\tilde{\boldsymbol{W}}_{lP}}{p_{\mathrm{p}}M} \to 0$, as $M \to \infty$. Therefore, (23) becomes

$$\frac{1}{\sqrt{p_{\mathrm{u}}}}\boldsymbol{r}_l \to \boldsymbol{D}_{ll}^{-1}\boldsymbol{D}_l\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^{\dagger}\frac{\boldsymbol{A}^{\dagger}\boldsymbol{A}}{M}\sum_{j=1}^{L}\boldsymbol{G}_{jl}\right)^{-1}\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^{\dagger}\right)\frac{\boldsymbol{A}^{\dagger}\boldsymbol{A}}{M}\left(\sum_{j=1}^{L}\boldsymbol{G}_{jl}\boldsymbol{x}_j\right). \tag{25}$$

As in the MRC case, for a ZF receiver, when the number of BS antennas grows without bound, the effect of noise vanishes. However, the pilot contamination effect persists even under a finite-dimensional channel model.

4.1.3 Uniform Linear Array

We now consider the special case of a uniform linear array. Here the response vector is given by

$$\boldsymbol{a}\left(\phi_p\right) = \frac{1}{\sqrt{P}}\left[1, e^{-j2\pi\frac{d}{\lambda}\sin\phi_p}, ..., e^{-j2\pi\frac{(M-1)d}{\lambda}\sin\phi_p}\right]^T, \tag{26}$$

where d is the antenna spacing, and λ is the carrier wavelength. For $p \neq q$, we have

$$\begin{aligned}\frac{1}{M}\boldsymbol{a}^\dagger\left(\phi_p\right)\boldsymbol{a}\left(\phi_q\right) &= \frac{1}{MP}\sum_{m=0}^{M-1} e^{j2\pi\frac{d}{\lambda}(\sin\phi_p-\sin\phi_q)m} \\ &= \frac{1}{MP}\frac{1-e^{j2\pi\frac{d}{\lambda}(\sin\phi_p-\sin\phi_q)M}}{1-e^{j2\pi\frac{d}{\lambda}(\sin\phi_p-\sin\phi_q)}} \overset{M\to\infty}{\to} 0.\end{aligned} \tag{27}$$

For $p = q$, $\sin\phi_p = \sin\phi_q$, so $\frac{1}{M}\boldsymbol{a}^\dagger\left(\phi_p\right)\boldsymbol{a}\left(\phi_q\right) = \frac{1}{P}$. Therefore,

$$\frac{1}{M}\boldsymbol{A}^\dagger\boldsymbol{A} \overset{M\to\infty}{\to} \frac{1}{P}\boldsymbol{I}_P. \tag{28}$$

Substitutions of (28) into (22) and (25) yield

$$\frac{1}{\sqrt{p_\mathrm{u}}M}\boldsymbol{r}_l \overset{M\to\infty}{\to} \frac{1}{P}\boldsymbol{D}_{ll}\boldsymbol{D}_l^{-1}\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^\dagger\right)\left(\sum_{j=1}^{L}\boldsymbol{G}_{jl}\boldsymbol{x}_j\right), \quad \text{for MRC} \tag{29}$$

$$\frac{1}{\sqrt{p_\mathrm{u}}}\boldsymbol{r}_l \overset{M\to\infty}{\to} \boldsymbol{D}_{ll}^{-1}\boldsymbol{D}_l\left[\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^\dagger\right)\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}\right)\right]^{-1}\left(\sum_{i=1}^{L}\boldsymbol{G}_{il}^\dagger\right)\left(\sum_{j=1}^{L}\boldsymbol{G}_{jl}\boldsymbol{x}_j\right), \quad \text{for ZF.} \tag{30}$$

Since the elements of $\boldsymbol{G}_{il}$ are independent, the above results reveal that the performance of the system under the finite-dimensional channel model with P angular bins and with an unlimited number of BS antennas is the same as the performance under an uncorrelated channel model with P antennas.

Remark 6 *Consider the case when $M, P \to \infty$ and assume that M grows at a greater rate than P. Then $\frac{1}{P}\boldsymbol{G}_{il}^\dagger\boldsymbol{G}_{jl} \to \delta_{ij}\boldsymbol{D}_{il}^{1/2}\boldsymbol{I}_K\boldsymbol{D}_{jl}^{1/2}$, where δ_{ij} is the delta function. Then, from* (29) *and* (30)*, we have*

$$\lim_{M,P\to\infty}\boldsymbol{C}_l\boldsymbol{r}_l \to \sum_{i=1}^{L}\boldsymbol{D}_{il}\boldsymbol{x}_l, \tag{31}$$

where $\boldsymbol{C}_l$ equals $\frac{1}{\sqrt{p_\mathrm{u}}M}\boldsymbol{D}_l\boldsymbol{D}_{ll}^{-1}$ for the MRC case and $\frac{1}{\sqrt{p_\mathrm{u}}}\boldsymbol{D}_{ll}$ for the ZF case. The effective signal-to-interference ratios (SIR) of the uplink transmission from the kth

user in the lth cell to the lth BS for the MRC and ZF receivers are thus equal, and given by

$$\mathsf{SIR}_{lk} = \frac{\beta_{llk}^2}{\sum_{i\neq l}^{L} \beta_{ilk}^2}. \tag{32}$$

The SIR in (32) *is equal to the SIR obtained in [8] which assumes the channel vectors for different users are asymptotically orthogonal, and that the BS uses MRC.*

4.2 Achievable Uplink Rates

In this subsection, we derive lower bounds on the achievable uplink rate for a finite (the analysis requires that $M \geq P$) and an infinite number of BS antennas both for the case of an MRC receiver and for a ZF receiver. To obtain these lower bounds we use the techniques of [21, 22], and use the channel estimate in (10). We assume a uniform array at the BS, and that the elements of $\boldsymbol{H}_{il}$, $i = 1, 2, .., L$, are i.i.d. Gaussian random variables. From (20), we have

$$\boldsymbol{r}_l = \boldsymbol{F}_{ll}^{\dagger} \left(\sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \hat{\boldsymbol{\Upsilon}}_{il} \boldsymbol{x}_i + \boldsymbol{n}_l - \sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \tilde{\boldsymbol{\Upsilon}}_{il} \boldsymbol{x}_i \right), \tag{33}$$

where $\tilde{\boldsymbol{\Upsilon}}_{il} \triangleq \hat{\boldsymbol{\Upsilon}}_{il} - \boldsymbol{\Upsilon}_{il}$. From the properties of MMSE estimation, $\tilde{\boldsymbol{\Upsilon}}_{il}$ is independent of $\hat{\boldsymbol{\Upsilon}}_{il}$. Let r_{lk} and x_{lk} be the kth elements of the $K \times 1$ vectors $\boldsymbol{r}_l$ and $\boldsymbol{x}_l$, respectively. Then,

$$\begin{aligned} r_{lk} &= \boldsymbol{F}_{llk}^{\dagger} \left(\sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \hat{\boldsymbol{\Upsilon}}_{il} \boldsymbol{x}_i + \boldsymbol{n}_l - \sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \tilde{\boldsymbol{\Upsilon}}_{il} \boldsymbol{x}_i \right) \\ &= \sqrt{p_{\mathrm{u}}} \boldsymbol{F}_{llk}^{\dagger} \hat{\boldsymbol{\Upsilon}}_{llk} x_{lk} + \mathcal{I}_{lk} - \sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \sum_{n=1}^{K} \boldsymbol{F}_{llk}^{\dagger} \tilde{\boldsymbol{\Upsilon}}_{iln} x_{in} + \boldsymbol{F}_{llk}^{\dagger} \boldsymbol{n}_l, \end{aligned} \tag{34}$$

where $\boldsymbol{F}_{llk}$, $\hat{\boldsymbol{\Upsilon}}_{ilk}$, and $\tilde{\boldsymbol{\Upsilon}}_{ilk}$ are the kth columns of the matrices $\boldsymbol{F}_{ll}$, $\hat{\boldsymbol{\Upsilon}}_{il}$, and $\tilde{\boldsymbol{\Upsilon}}_{il}$, respectively; and $\mathcal{I}_{lk} \triangleq \sqrt{p_{\mathrm{u}}} \sum_{n\neq k}^{K} \boldsymbol{F}_{llk}^{\dagger} \hat{\boldsymbol{\Upsilon}}_{lln} x_{ln} + \sqrt{p_{\mathrm{u}}} \sum_{i\neq l}^{L} \sum_{n=1}^{K} \boldsymbol{F}_{llk}^{\dagger} \hat{\boldsymbol{\Upsilon}}_{iln} x_{in}$. We present a lower bound on the achievable ergodic rate of the uplink transmission in the following proposition.

Proposition 9 *The uplink achievable ergodic rate of the kth user in the lth cell with Gaussian inputs for a linear detection matrix $\boldsymbol{F}_{ll}$ at the BS is lower bounded by*

$$R_{lk} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{p_{\mathrm{u}} \left| \boldsymbol{F}_{llk}^{\dagger} \hat{\boldsymbol{\Upsilon}}_{llk} \right|^2}{\mathbb{E}\left\{ |\mathcal{I}_{lk}|^2 | \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk} \right\} + p_{\mathrm{u}} \sum_{i=1}^{L} \sum_{n=1}^{K} \boldsymbol{F}_{llk}^{\dagger} \mathrm{cov}\left(\tilde{\boldsymbol{\Upsilon}}_{iln} \right) \boldsymbol{F}_{llk} + \| \boldsymbol{F}_{llk} \|^2} \right) \right\}, \tag{35}$$

where $\mathrm{cov}\,(\boldsymbol{x})$ *denotes the covariance matrix of the vector* $\boldsymbol{x}$.

Proof: *See Appendix A.* □

The above capacity lower bound can be approached if the message is encoded over many realizations of all sources of randomness that enter the model (noise and channel). In practice, assuming wideband operation, this can be achieved by coding over the frequency domain, using, for example coded OFDM. Note that if one were implementing a soft-decoder, say for LDPC error-correcting, then one would have to use an expression for the posterior likelihood of the coded bits; one would proceed exactly as we did in deriving the capacity lower bounds and make the same Gaussian approximation (which, in fact, is fairly accurate). Hence, our capacity lower bounds can be expected to describe rather accurately the performance of a soft-decoder (within a dB or so).

We next consider two specific linear detectors at the BS, namely, the MRC and ZF receivers.

4.2.1 Maximum-Ratio Combining

From Proposition 9, we obtain a lower bound on the achievable ergodic rate for the MRC technique as summarized in the following theorem.

Theorem 1 *A lower bound on the achievable ergodic rate of the kth user in the lth cell of the uplink transmission for the MRC performance at the BS is given by*

$$R_{lk}^{\mathrm{MRC}} = \mathbb{E}\left\{\log_2\left(1+\mathsf{SINR}_{lk}^{\mathrm{MRC}}\right)\right\}, \tag{36}$$

where

$$\mathsf{SINR}_{lk}^{\mathrm{MRC}} = \frac{p_{\mathrm{u}}\left\|\hat{\boldsymbol{\Upsilon}}_{llk}\right\|^4}{p_{\mathrm{u}}\frac{\sum_{i\neq l}^{L}\beta_{ilk}^2}{\beta_{llk}^2}\left\|\hat{\boldsymbol{\Upsilon}}_{llk}\right\|^4 + p_{\mathrm{u}}\hat{\boldsymbol{\Upsilon}}_{llk}^{\dagger}\sum\limits_{i=1}^{L}\left(\beta_{il}\boldsymbol{A}\boldsymbol{A}^{\dagger}-\mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{ilk}\right)\right)\hat{\boldsymbol{\Upsilon}}_{llk}+\left\|\hat{\boldsymbol{\Upsilon}}_{llk}\right\|^2}, \tag{37}$$

with $\mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{ilk}\right) = p_{\mathrm{p}}\beta_{ilk}^2\boldsymbol{A}\left(p_{\mathrm{p}}\boldsymbol{A}^{\dagger}\boldsymbol{A}\sum_{j=1}^{L}\beta_{jlk}+\boldsymbol{I}_P\right)^{-1}\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger}$, *and* $\beta_{il} \triangleq \sum_{k=1}^{K}\beta_{ilk}$.

Proof: *See Appendix B.* □

Corollary 1 *For an unlimited number of BS antennas, the uplink ergodic rate of the kth user in the lth cell achieved by using MRC is lower bounded by*

$$\tilde{R}_{lk}^{\mathrm{MRC}} = \frac{\tilde{\beta}_{lk}^{P}}{\beta_{llk}^{P}\,(P-1)!}\mathcal{K}_{P-1,\frac{\tilde{\beta}_{lk}}{\beta_{llk}}}\left(\beta_{llk}^{2}, \sum_{i\neq l}^{L}\beta_{ilk}^{2}, \sum_{i=1}^{L}\left(\beta_{il}-\frac{\beta_{ilk}^{2}}{\tilde{\beta}_{lk}}\right)\beta_{llk}\right), \tag{38}$$

where $\tilde{\beta}_{lk} \triangleq \sum_{i=1}^{L}\beta_{ilk}$*, and* $\mathcal{K}_{n,\mu}(a,b,c)$ *is defined as*

$$\begin{aligned}\mathcal{K}_{n,\mu}(a,b,c) \triangleq \frac{\log_2 e}{\mu^{n+1}}\sum_{i=0}^{n}\frac{n!}{(n-i)!}\Bigg[&-\left(\frac{-\mu c}{a+b}\right)^{n-i}e^{\frac{\mu c}{a+b}}\mathrm{Ei}\left(\frac{-\mu c}{a+b}\right)\\ &+\left(\frac{-\mu c}{b}\right)^{n-i}e^{\frac{\mu c}{b}}\mathrm{Ei}\left(\frac{-\mu c}{b}\right)+\sum_{k=1}^{n-i}(k-1)!\left(\left(\frac{-\mu c}{a+b}\right)^{n-i-k}-\left(\frac{-\mu c}{b}\right)^{n-i-k}\right)\Bigg],\end{aligned} \tag{39}$$

and where $\mathrm{Ei}(\cdot)$ *is the exponential integral function.*

Proof: *See Appendix C.* □

4.2.2 Zero-Forcing Receiver

From Proposition 9, we obtain a lower bound on the achievable ergodic rate for the ZF receiver as stated in the following theorem.

Theorem 2 *A lower bound on the achievable ergodic rate of the kth user in the lth cell of the uplink transmission assuming ZF processing at the BS is given by*

$$R_{lk}^{\mathrm{ZF}} = \mathbb{E}\left\{\log_2\left(1+\mathsf{SINR}_{lk}^{\mathrm{ZF}}\right)\right\}, \tag{40}$$

where

$$\mathsf{SINR}_{lk}^{\mathrm{ZF}} = \frac{p_{\mathrm{u}}}{p_{\mathrm{u}}\frac{\sum_{i\neq l}^{L}\beta_{ilk}^{2}}{\beta_{ilk}^{2}}+p_{\mathrm{u}}\sum_{i=1}^{L}\sum_{n=1}^{K}\left[\boldsymbol{\Xi}^{-1}\hat{\boldsymbol{\Upsilon}}_{ll}^{\dagger}\mathrm{cov}\left(\tilde{\boldsymbol{\Upsilon}}_{iln}\right)\hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{\Xi}^{-1}\right]_{kk}+\left[\boldsymbol{\Xi}^{-1}\right]_{kk}}, \tag{41}$$

with $\boldsymbol{\Xi} \triangleq \hat{\boldsymbol{\Upsilon}}_{ll}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{ll}$*, and*

$$\mathrm{cov}\left(\tilde{\boldsymbol{\Upsilon}}_{iln}\right) = \beta_{iln}\boldsymbol{A}\boldsymbol{A}^{\dagger} - p_{\mathrm{p}}\beta_{iln}^{2}\boldsymbol{A}\left(p_{\mathrm{p}}\boldsymbol{A}^{\dagger}\boldsymbol{A}\tilde{\beta}_{ln}+\boldsymbol{I}_{P}\right)^{-1}\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger}.$$

Proof: *For the ZF receiver,* $\boldsymbol{F}_{ll} = \hat{\boldsymbol{\Upsilon}}_{ll}\left(\hat{\boldsymbol{\Upsilon}}_{ll}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{ll}\right)^{-1}$*. Since* $\hat{\boldsymbol{\Upsilon}}_{il} = \hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{D}_{ll}^{-1}\boldsymbol{D}_{il}$*,*

$$\boldsymbol{F}_{ll}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{il} = \left(\hat{\boldsymbol{\Upsilon}}_{ll}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{ll}\right)^{-1}\hat{\boldsymbol{\Upsilon}}_{ll}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{D}_{ll}^{-1}\boldsymbol{D}_{il} = \boldsymbol{D}_{ll}^{-1}\boldsymbol{D}_{il},$$

leading to

$$\boldsymbol{F}_{llk}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{iln} = \delta_{kn}\frac{\beta_{iln}}{\beta_{llk}}, \tag{42}$$

where δ_{kn} is the delta function. Substituting (42) *into* (35)*, we obtain* (40). □

Corollary 2 *For an unlimited number of BS antennas, the uplink ergodic rate of the kth user in the lth cell achieved by using ZF is lower bounded by*

$$\tilde{R}_{lk}^{\mathrm{ZF}} = \frac{\tilde{\beta}_{lk}^{P-K+1}}{\beta_{llk}^{2(P-K+1)}(P-K)!}\mathcal{K}_{P-K,\frac{\tilde{\beta}_{lk}}{\beta_{llk}^2}}\left(1, \frac{\sum_{i\neq l}^{L}\beta_{ilk}^2}{\beta_{llk}^2}, \sum_{i=1}^{L}\sum_{n=1}^{K}\frac{\beta_{iln}\sum_{j\neq i}^{L}\beta_{jln}}{\tilde{\beta}_{ln}}\right). \tag{43}$$

Proof: *See Appendix D.* □

Remark 7 *Consider* (58) *and* (66) *in Appendices C and D, when both the number of BS antennas M and the number of physical directions P grow without bound (assuming that M grows at a greater rate than P). Then, the lower bounds on the uplink rates for MRC and ZF receivers (i.e., R_{lk}^{MRC} and R_{lk}^{ZF}, respectively) approach the same rate R_{lk}^{∞} given by*

$$R_{lk}^{\infty} = \log_2\left(1 + \frac{\beta_{llk}^2}{\sum_{i\neq l}^{L}\beta_{ilk}^2}\right), \tag{44}$$

which equals the asymptotic rate corresponding to the SINR in (32)*. This is due to the fact that when M and P are large, things that were random before become deterministic and hence, the lower bound approaches a non-random value.*

5 Numerical Results

In this section, we give some numerical results to verify our analysis. We first consider a simple scenario to study the fundamental effects of pilot contamination, the number of BS antennas, and the number of physical directions (dimension of the channel model) on the system performance for MRC and ZF receivers. Then, we consider a more practical scenario, which is similar to the simulation model used in [8], in order to further compare the performances of the MRC and ZF receivers for an unlimited number of BS antennas. In all examples, we choose a uniform linear array at the BS with a relative element spacing of $\frac{d}{\lambda} = 0.3$ and uniformly distributed arrival angles $\phi_p = -\pi/2 + (p-1)\,\pi/P$, for $p = 1, 2, ..., P$.

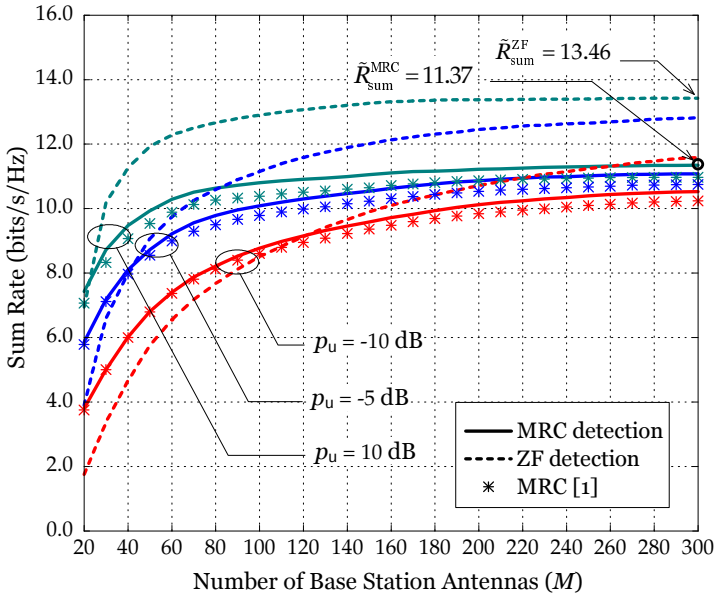


Figure 2: Lower bound on the uplink sum-rate versus the number of BS antennas, for $a = 0.1$ and $P = 20$.

5.1 Scenario I

We consider a system with 4 cells. Each cell has a BS that serves $K = 10$ users. The training sequence is $\tau = K$ symbols long. (With a training sequence of length of $\tau = K$, the BS can learn the channels of K users.) We assume that all direct gains are equal to 1 and that all cross-gains are equal to a, i.e, $\beta_{llk} = 1$ and $\beta_{ilk} = a$ $\forall i \neq l$, $k = 1, ..., K$. We define the following lower bounds on the uplink sum-rates in the lth cell for the MRC and ZF schemes:

$$R_{\text{sum}}^{\text{MRC}} \triangleq \sum_{k=1}^{K} R_{lk}^{\text{MRC}}, \; R_{\text{sum}}^{\text{ZF}} \triangleq \sum_{k=1}^{K} R_{lk}^{\text{ZF}}.$$

To study the effect of the number of BS antennas on the system performance, Fig. 2 shows the lower bounds on the uplink sum-rates in the lth cell versus the number of antennas M, for $a = 0.1$, $P = 20$, for different average transmit powers per user $p_{\text{u}} = -10, -5$, and 10 dB.[1] The "MRC [1]" curves represent the bounds for MRC obtained from [23]. Clearly, our new bound is tighter than the one in [23]. We can see that using a large number of BS antennas significantly improves the achievable rate. However, when the number of BS antennas increases beyond a certain point (for example $M = 80$ for $p_{\text{u}} = 10$ dB and $M = 140$ for $p_{\text{u}} = -5$ dB), the sum-rates

[1]Since the noise power is unity, the SNR is equal to p_{u}.

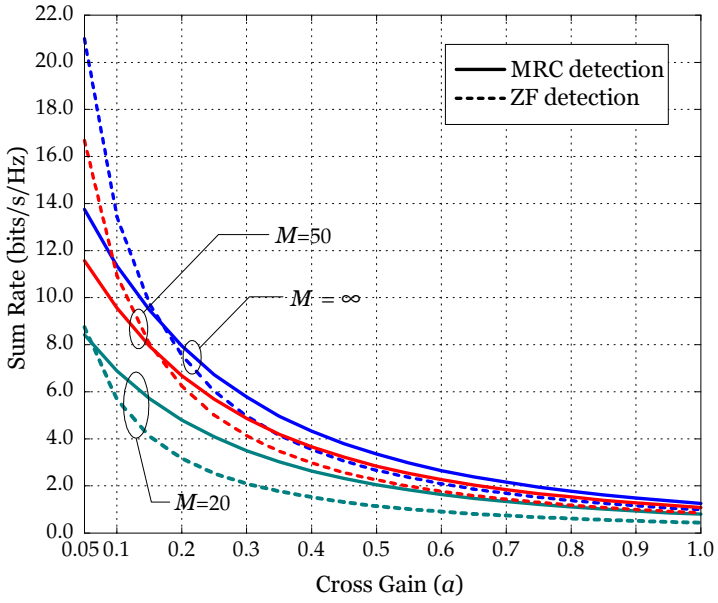


Figure 3: Lower bound on the uplink sum-rate versus the cross gain, for $p_{\mathrm{u}} = 0$ dB and $P = 20$.

increase very slowly. This means that a large but finite number of antennas can offer a performance which is close to the performance that could ultimately be achieved with an unlimited number of antennas. We can see that ZF is preferable to MRC at high SNR, and when adding more BS antennas the performance advantage of ZF over MRC widens. This is due to the fact that when the number of BS antennas increases, the SINR increases, while the MRC technique works well at low SNR and ZF is better at high SNR. Furthermore, we can see that as $M \to \infty$, the sum-rates approach $\tilde{R}_{\mathrm{sum}}^{\mathrm{MRC}}$ and $\tilde{R}_{\mathrm{sum}}^{\mathrm{ZF}}$ for the MRC and ZF schemes, respectively. These are the asymptotic sum-rates with an unlimited number of BS antennas and they are independent of the SNR (cf. (38) and (43)).

We now consider the effect of pilot contamination. Fig. 3 depicts the lower bounds on the sum-rates for the uplink transmission versus the cross gain at $p_{\mathrm{u}} = 0$ dB and $P = 20$ for different number of BS antennas: $M = 20$, 50 and ∞. We can see that the effect of pilot contamination can be very significant if the value of the cross gain is close to the value of the direct gain, regardless of how many antennas the BS is equipped with. Furthermore, for low cross-gain values, using the ZF receiver can offer better sum-rate performance compared with the MRC technique, and vice versa for high cross gain values. This is again due to the fact that the ZF receiver works best at high signal-to-interference-plus-noise ratio (SINR). For high values of a, the SINR decreases due to the pilot contamination effect and hence, the system performance with ZF significantly degrades.

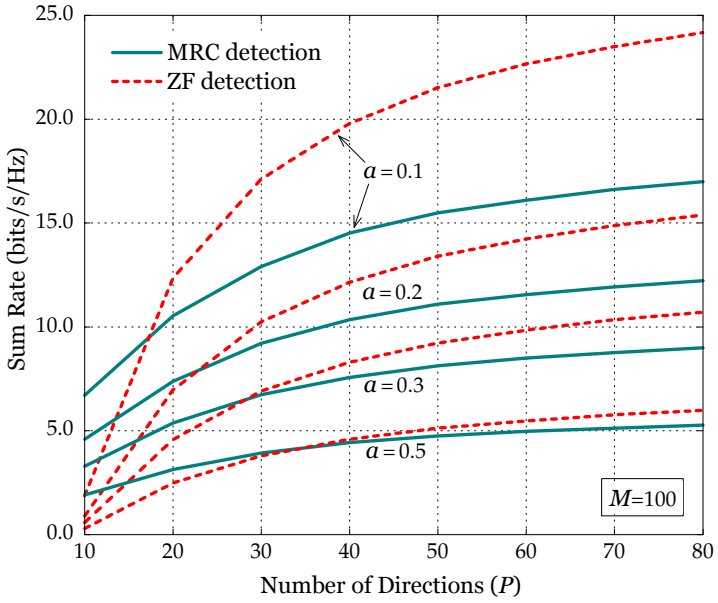


Figure 4: Lower bound on the uplink sum-rate versus the number of directions P, for $M = 100$ and $p_{\mathrm{u}} = 0$ dB.

Fig. 4 shows the lower bounds on the uplink sum-rates versus the number of directions P, at $p_{\mathrm{u}} = 0$ dB, $M = 100$, for $a = 0.1$, $a = 0.2$, $a = 0.3$, and $a = 0.5$. We can see that the sum-rate increases with increasing P. In particular, the ZF scheme yields better performance than the MRC scheme in rich scattering propagation environments even when the cross gain is large and that the pilot contamination effect is substantial. The fact that ZF is preferable over MRC in rich scattering propagation environments is further illustrated in Fig. 2 which shows the lower bounds on the uplink sum-rates versus the cross gain a for an unlimited number of BS antennas and for $P = 20$, 50, 100, 500, and ∞.

5.2 Scenario II

We consider a hexagonal cellular network, similar to the one discussed in [8]. Each cell has a radius (from center to vertex) of r_{c}. The number of users per cell is $K = 10$ and we assume that no user is closer to the BS than $r_{\mathrm{c}}^{'}$ meters. The BS has an infinite number of antennas. We assume that the transmitted data are modulated with OFDM with an OFDM symbol duration of T_{s}. The useful symbol duration is T_{u}, and the cyclic prefix interval is $T_{\mathrm{g}} = T_{\mathrm{s}} - T_{\mathrm{u}}$. Let T_{slot} be the slot length, and let T_{pilot} be the time used for the transmission of pilots. Hence, the time spent on the data transmission is $T_{\mathrm{slot}} - T_{\mathrm{pilot}}$. Therefore, following [8], we

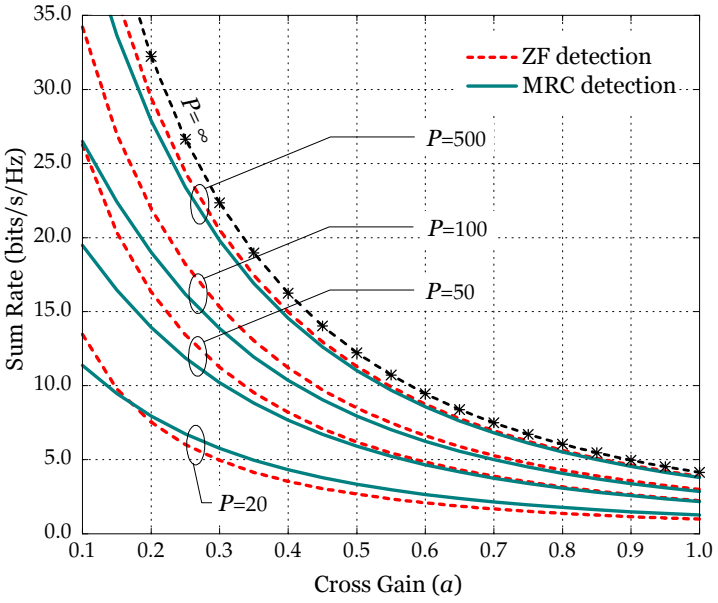


Figure 5: Lower bound on the uplink sum-rate versus the cross gain a, for $M = \infty$.

define the following lower bound on the *net* uplink rate of the kth user in the lth cell:

$$R_{lk}^{\text{net}} = \left(\frac{B}{\alpha}\right)\left(\frac{T_{\text{slot}} - T_{\text{pilot}}}{T_{\text{slot}}}\right)\left(\frac{T_{\text{u}}}{T_{\text{s}}}\right)\tilde{R}_{lk} \quad \text{bits/sec.}$$

where $\tilde{R}_{lk}$ equals $\tilde{R}_{lk}^{\text{MRC}}$ for MRC and $\tilde{R}_{lk}^{\text{ZF}}$ for ZF, B is the total bandwidth, and α is the frequency reuse factor. Here, $(T_{\text{slot}} - T_{\text{pilot}})/T_{\text{slot}}$ reflects the pilot overhead, i.e., the ratio between the time used for data transmission and the total slot length. Also, $T_{\text{u}}/T_{\text{s}}$ represents the overhead incurred by the cyclic prefix. For the simulation, we choose parameters that resemble those of the LTE standard: $T_{\text{s}} = 71.4$ μsec, $T_{\text{u}} = 66.7$ μsec, a subcarrier spacing of $\Delta_f = 15$ KHz. We choose the channel coherence time to be 500 μsec (this is equivalent to $500/71.4 \approx 7$ OFDM symbols). The BS is serving 10 users, so one OFDM symbol is used for uplink pilots (with one symbol, the BS can learn the channel for a maximum of $\frac{1}{T_{\text{g}}\Delta_f} \approx 14$ users), one symbol is used for the additional overhead, and the remaining five symbols are spent on payload data. Therefore, $(T_{\text{slot}} - T_{\text{pilot}})/T_{\text{slot}} = 5/7$. We assume a total system bandwidth of $B = 20$ MHz. Furthermore, the large-scale fading β_{ilk} is modeled via z_{ilk}/r_{ilk}^{γ}, where z_{ilk} is a log-normal random variable with standard deviation σ_{shadow}, where r_{ilk} is the distance between the kth user in the ith cell and the lth BS, and γ is the path loss exponent. We assume that the users are randomly located in each cell.

Fig. 5 shows the cumulative distribution of the lower bound on the net uplink rate per user for different reuse factors $\alpha = 1$, 3, and 7. Results are shown for the

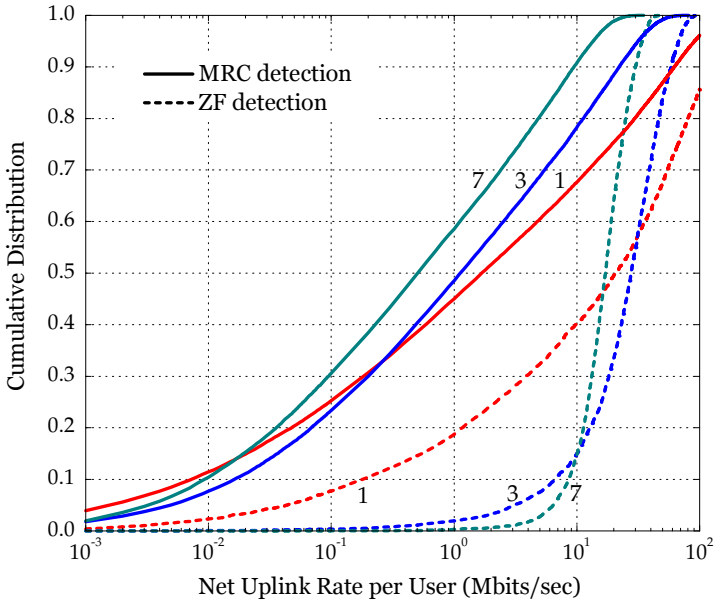


Figure 6: Cumulative distribution of the lower bound on the net uplink rate per user, for frequency reuse factors 1, 3, and 7. Here $M = \infty$, $P = 50$, $r_c = 1600$ m, $\sigma_{\text{shadow}} = 8$ dB, and $\gamma = 3.8$.

MRC and the ZF receivers. Here, we choose $r_c = 1600$ m, $r_c' = 100$ m, $\gamma = 3.8$, $\sigma_{\text{shadow}} = 8$ dB, and $P = 50$. The ZF scheme outperforms the MRC scheme in this example, and the distribution of the net uplink rate for the ZF scheme is more concentrated around its median compared to the MRC scheme. Furthermore, at high SIR (and hence at high rate), smaller reuse factors are preferable, and vice versa at low SIR. Table 1 summarizes the 95%-likely net uplink rates.

We next consider $r_c = 1000$ m, $r_c' = 100$ m, $\gamma = 2.2$, $\sigma_{\text{shadow}} = 8$ dB, and $P = 15$. The cumulative distributions of the lower rate bounds for the MRC and ZF receivers with different reuse factors $\alpha = 1$, 3, and 7 are plotted in Fig. 7. Comparing with the setting in Fig. 5, here we reduce the cell radius, and the path loss exponent, so the effect of pilot contamination will increase. Furthermore, we consider the performance for small P. It can be seen from the figure that MRC yields better performance than ZF for a reuse factor of 1. However, for reuse factors 3 or 7 (where there is less pilot contamination), the ZF technique is better. This means that when the effect of pilot contamination is large and P is small, MRC is preferable over ZF. These conclusions are the same as those drawn in Scenario I. Table 2 summarizes the uplink performance for this setting.

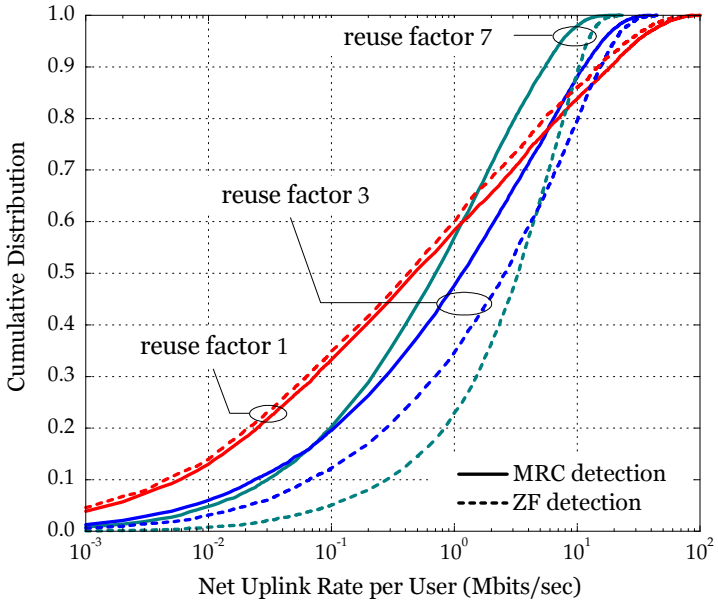


Figure 7: Same as Figure 5, but with $P = 15$, $r_c = 1000$ m and $\gamma = 2.2$.

Table 1: Uplink performance of MRC and ZF with frequency reuse factors 1, 3, and 7, for $M = \infty$, $P = 50$, $r_c = 1600$ m, $\sigma_{\text{shadow}} = 8$ dB, and $\gamma = 3.8$.

Frequency Reuse Factor	95%-likely Net Uplink Rate (Mbits/sec)		Mean of the Net Uplink Rate per User (Mbits/sec)	
	MRC	ZF	MRC	ZF
1	0.002	0.04	17.5	41.4
3	0.005	3.10	6.5	30.4
7	0.003	6.42	2.8	18.2

6 Conclusions

This paper has analyzed the pilot contamination effect in multicell MU-MIMO systems with very large antenna arrays at the BS. In particular, we have studied a model for the physical channel where a *finite* number of scattering centers P are visible from the BS. We showed that the pilot contamination effect discovered in [8] persists under this finite-dimensional channel model. An infinite number of antennas enables the receiver to access the signal arriving from each of the P scatterers, and in effect we have a spatially distributed receive array comprising antennas located at the positions of the P scatterers.

Furthermore, we derived a lower bound on the achievable uplink ergodic rate using linear detection at the BS. We deduced specific lower bounds on this rate for

Table 2: Same as Table 1, but with $P = 15$, $r_c = 1000$ m and $\gamma = 2.2$.

Frequency Reuse Factor	95%-likely Net Uplink Rate (Kbits/sec)		Mean of the Net Uplink Rate per User (Mbits/sec)	
	MRC	ZF	MRC	ZF
1	1.5	1.2	5.5	4.6
3	7.6	21	3.7	5.5
7	10	97	1.8	4.4

the cases of MRC and ZF receivers. We found that the ZF receiver can offer a higher sum-rate compared to MRC, when the pilot contamination effect is low, and vice versa. Furthermore, we observed that ZF becomes increasingly beneficial when adding more BS antennas. We have also found that the system performances with MRC and ZF depend on the number of physical directions P. ZF is preferable for large P (i.e., in a rich scattering propagation environment), while MRC is better for small P. The radically different qualitative behavior which was observed as we changed the propagation parameters implies that large-scale propagation experiments are urgently needed.

Appendix

A Proof of Proposition 9

The achievable rate of the kth user in the lth cell for the uplink transmission when we use the linear detection matrix $\boldsymbol{F}_{ll}$ at the BS is given by the mutual information between the unknown transmitted signal x_{lk} and the observed received signal r_{lk} and the known channel estimate $\hat{\boldsymbol{\Upsilon}}_{ll}$, i.e., $I\left(x_{lk}; r_{lk}, \hat{\boldsymbol{\Upsilon}}_{ll}\right)$. Using a similar approach as in [21, 22], we have

$$I\left(x_{lk}; r_{lk}, \hat{\boldsymbol{\Upsilon}}_{ll}\right) \geq I\left(x_{lk}; r_{lk}, \boldsymbol{F}_{ll}, \hat{\boldsymbol{\Upsilon}}_{llk}\right) \geq I\left(x_{lk}; r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right), \tag{45}$$

where the above inequalities follow from the fact that the mutual information cannot increase by performing additional processing operations on $\hat{\boldsymbol{\Upsilon}}_{ll}$ and $\boldsymbol{F}_{ll}$ [24]. Assuming x_{lk} is Gaussian distributed with unit variance, we have

$$\begin{aligned} I\left(x_{lk}; r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right) &= h\left(x_{lk}\right) - h\left(x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right) \\ &= \log_2\left(\pi e\right) - h\left(x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right). \end{aligned} \tag{46}$$

Since the differential entropy of a random variable with fixed variance is maximized when the random variable is Gaussian, we have

$$I\left(x_{lk}; r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right) \geq \log_2\left(\pi e\right) - \mathbb{E}\left\{\log_2\left(\pi e \text{ var}\left(x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right)\right)\right\}, \tag{47}$$

where $\text{var}\left(x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right) = \mathbb{E}\left\{\left|x_{lk} - \mathbb{E}\left\{x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right\}\right|^2\right\}$. The bound in (47) is tighter when $\text{var}\left(x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right)$ attains it minimum, which means that the value $\mathbb{E}\left\{x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right\}$ is the linear minimum mean-square-error estimate of x_{lk}, leading to

$$\text{var}\left(x_{lk}|r_{lk}, \boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right) = 1 - \frac{p_{\text{u}}\left|\boldsymbol{F}_{llk}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{llk}\right|^2}{\mathbb{E}\left\{\left|r_{lk}\right|^2 |\boldsymbol{F}_{llk}, \hat{\boldsymbol{\Upsilon}}_{llk}\right\}}. \tag{48}$$

From (3), (5), and (10), we have $\hat{\boldsymbol{\Upsilon}}_{il} = \hat{\boldsymbol{\Upsilon}}_{ll}\boldsymbol{D}_{ll}^{-1}\boldsymbol{D}_{il}$. Since $\tilde{\boldsymbol{\Upsilon}}_{il}$ and $\hat{\boldsymbol{\Upsilon}}_{il}$ are uncorrelated, $\tilde{\boldsymbol{\Upsilon}}_{il}$ and $\hat{\boldsymbol{\Upsilon}}_{ll}$ are uncorrelated for all i and hence, $\boldsymbol{F}_{llk}$ and $\hat{\boldsymbol{\Upsilon}}_{llk}$ are uncorrelated with $\tilde{\boldsymbol{\Upsilon}}_{iln}$ for all i and n (since $\boldsymbol{F}_{llk}$ is a function of $\hat{\boldsymbol{\Upsilon}}_{ll}$). Furthermore, $\boldsymbol{F}_{llk}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{llk}x_{lk}$, $\mathcal{I}_{lk}$, $\boldsymbol{F}_{llk}^{\dagger}\tilde{\boldsymbol{\Upsilon}}_{iln}x_{in}$, and $\boldsymbol{F}_{llk}^{\dagger}\boldsymbol{n}_l$ are uncorrelated. Therefore from (34), we obtain

$$\mathbb{E}\left\{|r_{lk}|^2\,|\boldsymbol{F}_{llk},\hat{\boldsymbol{\Upsilon}}_{llk}\right\} = p_{\mathrm{u}}\left|\boldsymbol{F}_{llk}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{llk}\right|^2 + \mathbb{E}\left\{|\mathcal{I}_{lk}|^2|\boldsymbol{F}_{llk},\hat{\boldsymbol{\Upsilon}}_{llk}\right\} + \sqrt{p_{\mathrm{u}}}\sum_{i=1}^{L}\sum_{n=1}^{K}\boldsymbol{F}_{llk}^{\dagger}\mathrm{cov}\left(\tilde{\boldsymbol{\Upsilon}}_{iln}\right)\boldsymbol{F}_{llk} + \|\boldsymbol{F}_{llk}\|^2. \quad (49)$$

Substituting (48) and (49) into (47), we obtain the lower bound on the uplink achievable ergodic rate of the kth user in the lth cell stated in Proposition 9.

B Proof of Theorem 1

For the MRC technique, $\boldsymbol{F}_{ll} = \hat{\boldsymbol{\Upsilon}}_{ll}$. Then, we have

$$\mathcal{I}_{lk} = \sqrt{p_{\mathrm{u}}}\sum_{n\neq k}^{K}\hat{\boldsymbol{\Upsilon}}_{llk}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{lln}x_{ln} + \sqrt{p_{\mathrm{u}}}\sum_{i\neq l}^{L}\sum_{n=1}^{K}\hat{\boldsymbol{\Upsilon}}_{llk}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{iln}x_{in}. \quad (50)$$

From (10) and $\hat{\boldsymbol{\Upsilon}}_{ilk} = \sqrt{\beta_{ilk}}\boldsymbol{A}\hat{\boldsymbol{h}}_{ilk}$, we have $\hat{\boldsymbol{\Upsilon}}_{ilk} = \frac{\beta_{ilk}}{\beta_{llk}}\hat{\boldsymbol{\Upsilon}}_{llk}$, thus

$$\mathbb{E}\left\{|\mathcal{I}_{lk}|^2\,|\boldsymbol{F}_{llk},\hat{\boldsymbol{\Upsilon}}_{llk}\right\} = \mathbb{E}\left\{|\mathcal{I}_{lk}|^2\,|\hat{\boldsymbol{\Upsilon}}_{llk}\right\} = p_{\mathrm{u}}\frac{\sum_{i\neq l}^{L}\beta_{ilk}^2}{\beta_{llk}^2}\left\|\hat{\boldsymbol{\Upsilon}}_{llk}\right\|^4 + p_{\mathrm{u}}\sum_{i=1}^{L}\sum_{n\neq k}^{K}\hat{\boldsymbol{\Upsilon}}_{llk}^{\dagger}\mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{iln}\right)\hat{\boldsymbol{\Upsilon}}_{llk}. \quad (51)$$

We now find the covariance matrices of $\hat{\boldsymbol{\Upsilon}}_{iln}$ and $\tilde{\boldsymbol{\Upsilon}}_{iln}$. Since $\hat{\boldsymbol{\Upsilon}}_{iln} = \sqrt{\beta_{iln}}\boldsymbol{A}\hat{\boldsymbol{h}}_{iln}$, $\mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{iln}\right) = \beta_{iln}\boldsymbol{A}\mathrm{cov}\left(\hat{\boldsymbol{h}}_{iln}\right)\boldsymbol{A}^{\dagger}$. We first find the covariance matrix of $\hat{\boldsymbol{h}}_{iln}$. From (9), (10) and using the matrix inversion lemma, we obtain

$$\mathrm{cov}\left(\hat{\boldsymbol{h}}_{iln}\right) = p_{\mathrm{p}}\beta_{iln}\left(p_{\mathrm{p}}\boldsymbol{A}^{\dagger}\boldsymbol{A}\sum_{j=1}^{L}\beta_{jln} + \boldsymbol{I}_P\right)^{-1}\boldsymbol{A}^{\dagger}\boldsymbol{A}. \quad (52)$$

Then the covariance matrix of $\hat{\boldsymbol{\Upsilon}}_{iln}$ is given by

$$\mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{iln}\right) = p_{\mathrm{p}}\beta_{iln}^2\boldsymbol{A}\left(p_{\mathrm{p}}\boldsymbol{A}^{\dagger}\boldsymbol{A}\sum_{j=1}^{L}\beta_{jln} + \boldsymbol{I}_P\right)^{-1}\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger}. \quad (53)$$

Since $\tilde{\boldsymbol{\Upsilon}}_{iln} = \hat{\boldsymbol{\Upsilon}}_{iln} - \boldsymbol{\Upsilon}_{iln}$, and $\hat{\boldsymbol{\Upsilon}}_{iln}$ and $\tilde{\boldsymbol{\Upsilon}}_{iln}$ are uncorrelated, we have

$$\mathrm{cov}\left(\tilde{\boldsymbol{\Upsilon}}_{iln}\right) = \mathrm{cov}\left(\boldsymbol{\Upsilon}_{iln}\right) - \mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{iln}\right) = \beta_{iln}\boldsymbol{A}\boldsymbol{A}^{\dagger} - \mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{iln}\right). \quad (54)$$

Substituting (51), (54) into (35), and using (53), completes the proof of Theorem 1.

C Proof of Corollary 1

From (28), when then number of BS antennas M goes to infinity, we have

$$\frac{1}{M}\left\|\hat{\boldsymbol{\Upsilon}}_{llk}\right\|^2 = \beta_{llk}\hat{\boldsymbol{h}}_{llk}^{\dagger}\frac{\boldsymbol{A}^{\dagger}\boldsymbol{A}}{M}\hat{\boldsymbol{h}}_{llk} \to \frac{\beta_{llk}}{P}\left\|\hat{\boldsymbol{h}}_{llk}\right\|^2 \tag{55}$$

$$\frac{1}{M^2}\hat{\boldsymbol{\Upsilon}}_{llk}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger}\hat{\boldsymbol{\Upsilon}}_{llk} = \beta_{llk}\hat{\boldsymbol{h}}_{llk}^{\dagger}\frac{\left(\boldsymbol{A}^{\dagger}\boldsymbol{A}\right)^2}{M^2}\hat{\boldsymbol{h}}_{llk} \to \frac{\beta_{llk}}{P^2}\left\|\hat{\boldsymbol{h}}_{llk}\right\|^2, \tag{56}$$

and

$$\frac{1}{M^2}\hat{\boldsymbol{\Upsilon}}_{llk}^{\dagger}\mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{llk}\right)\hat{\boldsymbol{\Upsilon}}_{llk} \to \frac{\beta_{ilk}^2\beta_{llk}}{P^2\sum_{i=1}^{L}\beta_{ilk}}\left\|\hat{\boldsymbol{h}}_{llk}\right\|^2. \tag{57}$$

We divide the numerator and the denominator of $\mathsf{SINR}_{lk}^{\mathrm{MRC}}$ by M^2, and use (55), (56) and (57). We obtain

$$\tilde{R}_{lk}^{\mathrm{MRC}} = \mathbb{E}\left\{\log_2\left(1+\frac{\beta_{llk}^2\left\|\hat{\boldsymbol{h}}_{llk}\right\|^2}{\sum_{i\neq l}^{L}\beta_{ilk}^2\left\|\hat{\boldsymbol{h}}_{llk}\right\|^2+\sum_{i=1}^{L}\left(\sum_{n=1}^{K}\beta_{iln}\beta_{llk}-\frac{\beta_{ilk}^2\beta_{llk}}{\sum_{i=1}^{L}\beta_{ilk}}\right)}\right)\right\}. \tag{58}$$

From (52), for an unlimited number of BS antennas, the covariance matrix of $\hat{\boldsymbol{h}}_{llk}$ becomes $\frac{\beta_{llk}}{\sum_{i=1}^{L}\beta_{ilk}}\boldsymbol{I}_P$. Therefore $\frac{2\sum_{i=1}^{L}\beta_{ilk}}{\beta_{llk}}\left\|\hat{\boldsymbol{h}}_{llk}\right\|^2$ has a Chi-square distribution with $2P$ degrees of freedom. Let $X \triangleq \left\|\hat{\boldsymbol{h}}_{llk}\right\|^2$. The probability density function (PDF) of X is then given by

$$p_X(x) = \frac{\left(\sum_{i=1}^{L}\beta_{ilk}/\beta_{llk}\right)^P}{(P-1)!}x^{P-1}\exp\left(-\frac{\sum_{i=1}^{L}\beta_{ilk}}{\beta_{llk}}x\right), \quad x \geq 0. \tag{59}$$

We define

$$\mathcal{K}_{n,\mu}(a,b,c) \triangleq \int_0^{\infty}\log_2\left(1+\frac{ax}{bx+c}\right)x^n e^{-\mu x}dx \tag{60}$$

$$= \int_0^{\infty}\log_2\left(1+\frac{a+b}{c}x\right)x^n e^{-\mu x}dx - \int_0^{\infty}\log_2\left(1+\frac{b}{c}x\right)x^n e^{-\mu x}dx. \tag{61}$$

Using (58)–(60), and [25, eq. (4.337.5)] completes the proof.

D Proof of Corollary 2

From (28), when then number of BS antennas $M \to \infty$, we have

$$\left[\left(\hat{\boldsymbol{Y}}_{ll}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\right)^{-1}\right]_{kk} = \left[\left(\hat{\boldsymbol{G}}_{ll}^{\dagger}\boldsymbol{A}^{\dagger}\boldsymbol{A}\hat{\boldsymbol{G}}_{ll}\right)^{-1}\right]_{kk} = \frac{1}{M}\left[\left(\hat{\boldsymbol{G}}_{ll}^{\dagger}\frac{\boldsymbol{A}^{\dagger}\boldsymbol{A}}{M}\hat{\boldsymbol{G}}_{ll}\right)^{-1}\right]_{kk} \to 0, \quad (62)$$

$$\left(\hat{\boldsymbol{Y}}_{ll}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\right)^{-1}\hat{\boldsymbol{Y}}_{ll}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\left(\hat{\boldsymbol{Y}}_{ll}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\right)^{-1} \to \left(\hat{\boldsymbol{G}}_{ll}^{\dagger}\hat{\boldsymbol{G}}_{ll}\right)^{-1}, \quad (63)$$

$$\left(\hat{\boldsymbol{Y}}_{ll}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\right)^{-1}\hat{\boldsymbol{Y}}_{ll}^{\dagger}\mathrm{cov}\left(\hat{\boldsymbol{Y}}_{iln}\right)\hat{\boldsymbol{Y}}_{ll}\left(\hat{\boldsymbol{Y}}_{ll}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\right)^{-1} \to \frac{\beta_{iln}^2}{\sum_{j=1}^{L}\beta_{jln}}\left(\hat{\boldsymbol{G}}_{ll}^{\dagger}\hat{\boldsymbol{G}}_{ll}\right)^{-1}. \quad (64)$$

Since $\mathrm{cov}\left(\tilde{\boldsymbol{\Upsilon}}_{iln}\right) = \beta_{iln}\boldsymbol{A}\boldsymbol{A}^{\dagger} - \mathrm{cov}\left(\hat{\boldsymbol{\Upsilon}}_{iln}\right)$, using (63) and (64) we have

$$\left(\hat{\boldsymbol{Y}}_{ll}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\right)^{-1}\hat{\boldsymbol{Y}}_{ll}^{\dagger}\mathrm{cov}\left(\tilde{\boldsymbol{Y}}_{iln}\right)\hat{\boldsymbol{Y}}_{ll}\left(\hat{\boldsymbol{Y}}_{ll}^{\dagger}\hat{\boldsymbol{Y}}_{ll}\right)^{-1} \to \frac{\beta_{iln}\sum_{j\neq i}^{L}\beta_{jln}}{\sum_{j=1}^{L}\beta_{jln}}\left(\hat{\boldsymbol{G}}_{ll}^{\dagger}\hat{\boldsymbol{G}}_{ll}\right)^{-1}, \text{ as } M \to \infty. \quad (65)$$

Substituting (62) and (65) into (40), we obtain

$$\tilde{R}_{lk}^{\mathrm{ZF}} = \mathbb{E}\left\{\log_2\left(1+\frac{1}{\frac{\sum_{i\neq l}^{L}\beta_{ilk}^2}{\beta_{llk}^2}+\sum_{i=1}^{L}\sum_{n=1}^{K}\frac{\beta_{iln}\sum_{j\neq i}^{L}\beta_{jln}}{\tilde{\beta}_{ln}}\left[\left(\hat{\boldsymbol{G}}_{ll}^{\dagger}\hat{\boldsymbol{G}}_{ll}\right)^{-1}\right]_{kk}}\right)\right\}. \quad (66)$$

Since $\hat{\boldsymbol{G}}_{ll} = \hat{\boldsymbol{H}}_{ll}\boldsymbol{D}_{ll}^{1/2}$ and the covariance of $\hat{\boldsymbol{h}}_{llk}$ becomes $\frac{\beta_{llk}}{\sum_{i=1}^{L}\beta_{ilk}}\boldsymbol{I}_P$ as $M \to \infty$, $\hat{\boldsymbol{G}}_{ll}^{\dagger}\hat{\boldsymbol{G}}_{ll}$ is a central complex Wishart matrix with P degrees of freedom and covariance matrix $\boldsymbol{\Sigma}_{ll} = \mathrm{diag}\left\{\frac{\beta_{ll1}^2}{\tilde{\beta}_{l1}}, \frac{\beta_{ll2}^2}{\tilde{\beta}_{l2}}, ..., \frac{\beta_{llK}^2}{\tilde{\beta}_{lK}}\right\}$, i.e, $\hat{\boldsymbol{G}}_{ll}^{\dagger}\hat{\boldsymbol{G}}_{ll} \sim \tilde{\mathcal{W}}_K\left(P, \boldsymbol{\Sigma}_{ll}\right)$ [26]. Let $Y \triangleq 1/\left[\left(\hat{\boldsymbol{G}}_{ll}^{\dagger}\hat{\boldsymbol{G}}_{ll}\right)^{-1}\right]_{kk}$. Then Y has a complex central Wishart distribution, $Y \sim \tilde{\mathcal{W}}_1\left(P-K+1, \frac{\beta_{llk}^2}{\tilde{\beta}_{lk}}\right)$ [27]. The PDF of Y is thus given by

$$p_Y(y) = \frac{\tilde{\beta}_{lk}e^{-\tilde{\beta}_{lk}/\beta_{llk}^2 y}}{\beta_{llk}^2(P-K)!}\left(\frac{\tilde{\beta}_{lk}}{\beta_{llk}^2}y\right)^{P-K}, \quad y \geq 0. \quad (67)$$

From (66) and (67), we have

$$\tilde{R}_{lk}^{\mathrm{ZF}} = \int_0^{\infty}\log_2\left(1+\frac{y}{\frac{\sum_{i\neq l}^{L}\beta_{ilk}^2}{\beta_{llk}^2}y+\sum_{i=1}^{L}\sum_{n=1}^{K}\frac{\beta_{iln}\sum_{j\neq i}^{L}\beta_{jln}}{\tilde{\beta}_{ln}}}\right)p_Y(y)\,dy. \quad (68)$$

Using (60), we obtain the result in Corollary 2.

References

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

[2] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality?" *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

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

[4] T. L. Marzetta, "How much training is required for multiuser MIMO," in *Fortieth Asilomar Conference on Signals, Systems and Computers (ACSSC '06)*, Pacific Grove, CA, USA, Oct. 2006, pp. 359–363.

[5] J. Jose, A. Ashikhmin, P. Whitning, and S. Vishwanath, "Scheduling and preconditioning in multi-user MIMO TDD systems," in *Proc. IEEE Int. Conf. Communications Workshops (ICCW)*, Reno, Nevada, May 2008, pp. 4100–4105.

[6] K. Takeuchi, M. Vehkaperä, T. Tanaka, and R. R. Müller, "Replica analysis of general multiuser detection in MIMO DS-CDMA channels with imperfect CSI," in *Proc. IEEE International Symposium on Information Theory*, Toronto, Canada, Jul. 2008, pp. 514–518.

[7] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, Jun. 2009, pp. 2184–2188.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–46, Jan. 2013.

[10] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[11] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 24–39, 2013.

[12] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications*, CA, US, Jun. 2011, pp. 381–385.

[13] 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," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[14] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *Proc. 49th Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, Illinois, Sep. 2011.

[15] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 361–368, Apr. 2002.

[16] R. R. Müller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2495–2506, Sep. 2002.

[17] A. G. Burr, "Capacity bounds and estimates for the finite scatterers MIMO wireless channel," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 812–818, Jun. 2003.

[18] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[20] H. Cramér, *Random Variables and Probability Distributions*. Cambridge, UK: Cambridge University Press, 1970.

[21] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

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

[23] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3464–3467.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[25] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA: Academic, 2007.

[26] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

[27] D. A. Gore, R. W. Heath Jr., and A. J. Paulraj, "Transmit selection in spatial multiplexing systems," *IEEE Commun. Lett.*, vol. 6, no. 11, pp. 491–493, Nov. 2002.

Paper C

Aspects of Favorable Propagation in Massive MIMO

Refereed article published in proc. ESIPCO 2014.

Aspects of Favorable Propagation in Massive MIMO

Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta

Abstract

Favorable propagation, defined as mutual orthogonality among the vector-valued channels to the terminals, is one of the key properties of the radio channel that is exploited in Massive MIMO. However, there has been little work that studies this topic in detail. In this paper, we first show that favorable propagation offers the most desirable scenario in terms of maximizing the sum-capacity. One useful proxy for whether propagation is favorable or not is the channel condition number. However, this proxy is not good for the case where the norms of the channel vectors are not equal. For this case, to evaluate how favorable the propagation offered by the channel is, we propose a "distance from favorable propagation" measure, which is the gap between the sum-capacity and the maximum capacity obtained under favorable propagation. Secondly, we examine how favorable the channels can be for two extreme scenarios: i.i.d. Rayleigh fading and uniform random line-of-sight (UR-LoS). Both environments offer (nearly) favorable propagation. Furthermore, to analyze the UR-LoS model, we propose an urns-and-balls model. This model is simple and explains the singular value spread characteristic of the UR-LoS model well.

1 Introduction

Recently, there has been a great deal of interest in massive multiple-input multiple-output (MIMO) systems where a base station (BS) equipped with a few hundred antennas simultaneously serves several tens of terminals [1–3]. Such systems can deliver all the attractive benefits of traditional MIMO, but at a much larger scale. More precisely, massive MIMO systems can provide high throughput, communication reliability, and high power efficiency with linear processing [4].

One of the key assumptions exploited by massive MIMO is that the channel vectors between the BS and the terminals should be nearly orthogonal. This is called *favorable propagation*. With favorable propagation, linear processing can achieve optimal performance. More explicitly, on the uplink, with a simple linear detector such as the matched filter, noise and interference can be canceled out. On the downlink, with linear beamforming techniques, the BS can simultaneously beamform multiple data streams to multiple terminals without causing mutual interference. Favorable propagation of massive MIMO was discussed in the papers [4, 5]. There, the condition number of the channel matrix was used as a proxy for how favorable the channel is. These papers only considered the case that the channels are i.i.d. Rayleigh fading. However, in practice, owing to the fact that the terminals have different locations, the norms of the channels are not identical. As we will see here, in this case, the condition number is not a good proxy for whether or not we have favorable propagation.

In this paper, we investigate the favorable propagation condition of different channels. We first show that under favorable propagation, we maximize the sum-capacity under a power constraint. When the channel vectors are i.i.d., the singular value spread is a useful measure of how favorable the propagation environment is. However, when the channel vectors have different norms, this is not so. We also ask whether or not practical scenarios will lead to favorable propagation. To this end, we consider two extreme scenarios: i.i.d. Rayleigh fading and uniform random line-of-sight (UR-LoS). We show that both scenarios offer substantially favorable propagation. We also propose a simple urns-and-balls model to analyze the UR-LoS case. For the sake of the argument, we will consider the uplink of a single-cell system.

2 Single-Cell System Model

Consider the uplink of a single-cell system where K single-antenna terminals independently and simultaneously transmit data to the BS. The BS has M antennas and all K terminals share the same time-frequency resource. If the K terminals simultaneously transmit the K symbols $x_1, \ldots, x_K$, where $\mathbb{E}\,|x_k|^2 = 1$, then the $M \times 1$ received vector at the BS is

$$\boldsymbol{y} = \sqrt{\rho} \sum_{k=1}^{K} \boldsymbol{g}_k x_k + \boldsymbol{w} = \sqrt{\rho} \boldsymbol{G} \boldsymbol{x} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{x} = [x_1, \ldots, x_K]^T$, $\boldsymbol{G} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K]$, $\boldsymbol{g}_k \in \mathbb{C}^{M\times 1}$ is the channel vector between the BS and the kth terminal, and $\boldsymbol{w}$ is a noise vector. We assume that the elements of $\boldsymbol{w}$ are i.i.d. $CN(0,1)$ random variables (RVs). With this assumption, ρ has the interpretation of normalized "transmit" signal-to-noise ratio (SNR). The channel vector $\boldsymbol{g}_k$ incorporates the effects of large-scale fading and small-scale fading. More precisely, the mth element of $\boldsymbol{g}_k$ is modeled as:

$$g_k^m = \sqrt{\beta_k} h_k^m, \qquad k = 1, \ldots, K, \quad m = 1, \ldots, M, \tag{2}$$

where h_k^m is the small-scale fading and β_k represents the large-scale fading which depends on k but not on m. This assumption is reasonable if the distance between the BS antennas is much smaller than the distance between terminals and the BS. For example, with half-wavelength antenna spacing, at 2.6 GHz, a rectangular planar array has a physical size of only about 60×60 cm. By contrast, the distance between the terminals and the BS is typically hundreds of meters.

3 Favorable Propagation

To have favorable propagation, the channel vectors $\{\boldsymbol{g}_k\}$, $k = 1, \ldots, K$, should be pairwisely orthogonal. More precisely, we say that the channel offers *favorable propagation* if

$$\boldsymbol{g}_i^H \boldsymbol{g}_j = \begin{cases} 0, & i, j = 1, \ldots, K, \quad i \neq j \\ \|\boldsymbol{g}_k\|^2 \neq 0, & k = 1, \ldots, K. \end{cases} \tag{3}$$

In practice, the condition (3) will never be exactly satisfied, but (3) can be approximately achieved. For this case, we say that the channel offers *approximately favorable propagation*. Also, under some assumptions on the propagation environment, when M grows large and $k \neq j$, it holds that

$$\frac{1}{M} \boldsymbol{g}_k^H \boldsymbol{g}_j \to 0, \quad M \to \infty. \tag{4}$$

For this case, we say that the channel offers *asymptotically favorable propagation*.

The favorable propagation condition (3) does not offer only the optimal performance with linear processing but also represents the most desirable scenario from the perspective of maximizing the information rate. See the following section.

3.1 Favorable Propagation and Capacity

Consider the system model (1). We assume that the BS knows the channel $\boldsymbol{G}$. The sum-capacity is given by

$$C = \log_2 \left| \boldsymbol{I} + \rho \boldsymbol{G}^H \boldsymbol{G} \right|. \tag{5}$$

Next, we will show that, subject to a constraint on $\boldsymbol{G}$, under favorable propagation conditions (3), C achieves its largest possible value. Firstly, we assume $\{\|\boldsymbol{g}_k\|^2\}$ are given. For this case, by using the Hadamard inequality, we have

$$\begin{aligned} C =& \log_2 \left| \boldsymbol{I} + \rho \boldsymbol{G}^H \boldsymbol{G} \right| \leq \log_2 \left(\prod_{k=1}^{K} [\boldsymbol{I} + \rho \boldsymbol{G}^H \boldsymbol{G}]_{k,k} \right) \\ =& \sum_{k=1}^{K} \log_2 \left([\boldsymbol{I} + \rho \boldsymbol{G}^H \boldsymbol{G}]_{k,k} \right) = \sum_{k=1}^{K} \log_2 \left(1 + \rho \|\boldsymbol{g}_k\|^2 \right). \end{aligned} \tag{6}$$

We can see that the equality of (6) holds if and only if $\boldsymbol{G}^H \boldsymbol{G}$ is diagonal, so that (3) is satisfied. This means that, given a constraint on $\{\|\boldsymbol{g}_k\|^2\}$, the channel propagation with the condition (3) provides the maximum sum-capacity.

Secondly, we consider a more relaxed constraint on the channel $\boldsymbol{G}$: $\|\boldsymbol{G}\|_F^2$ is given. From (6), by using Jensen's inequality, we get

$$\begin{aligned} C \leq& \sum_{k=1}^{K} \log_2 \left(1 + \rho \|\boldsymbol{g}_k\|^2 \right) = K \cdot \frac{1}{K} \sum_{k=1}^{K} \log_2 \left(1 + \rho \|\boldsymbol{g}_k\|^2 \right) \\ \leq& K \log_2 \left(1 + \frac{\rho}{K} \sum_{k=1}^{K} \|\boldsymbol{g}_k\|^2 \right) = K \log_2 \left(1 + \frac{\rho}{K} \|\boldsymbol{G}\|_F^2 \right), \end{aligned} \tag{7}$$

where equality in the first step holds when (3) satisfied, and equality in the second step holds when all $\|\boldsymbol{g}_k\|^2$ are equal. So, for this case, C is maximized if (3) holds and $\{\boldsymbol{g}_k\}$ have the same norm. The constraint on $\boldsymbol{G}$ that results in (7) is more relaxed than the constraint on $\boldsymbol{G}$ that results in (6), but the bound in (7) is only tight if all $\{\boldsymbol{g}_k\}$ have the same norm.

3.2 Measures of Favorable Propagation

Clearly, to check whether the channel can offer favorable propagation or not, we can check directly the condition (3) or (4). Other simple methods to measure whether the channel offers favorable propagation is to consider the condition number, or the *distance from favorable propagation* (to be defined shortly). These measures will be discussed in more detail in the following subsections.

3.2.1 Condition Number

Under the favorable propagation condition (3), we have

$$\boldsymbol{G}^H \boldsymbol{G} = \mathrm{Diag}\{\|\boldsymbol{g}_1\|^2, \cdots, \|\boldsymbol{g}_K\|^2\}. \tag{8}$$

We can see that if $\{\boldsymbol{g}_k\}$ have the same norm, the condition number of the Gramian matrix $\boldsymbol{G}^H\boldsymbol{G}$ is equal to 1:

$$\sigma_{\max}/\sigma_{\min} = 1, \tag{9}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximal and minimal singular values of $\boldsymbol{G}^H\boldsymbol{G}$.

Similarly, if the channel offers asymptotically favorable propagation, then we have

$$\frac{1}{M}\boldsymbol{G}^H\boldsymbol{G} \to \boldsymbol{D}, \quad M \to \infty, \tag{10}$$

where $\boldsymbol{D}$ is a diagonal matrix whose kth diagonal element is β_k. So, if all $\{\beta_k\}$ are equal, then the condition number is asymptotically equal to 1.

Therefore, when the channel vectors have the same norm (the large scale fading coefficients are equal), we can use the condition number to determine how favorable the channel propagation is. Since the condition number is simple to evaluate, it has been used as a measure of how favorable the propagation offered by the channel $\boldsymbol{G}$ is, in the literature. However, it has two drawbacks: i) it only has a sound operational meaning when all $\{\boldsymbol{g}_k\}$ have the same norm or all $\{\beta_k\}$ are equal; and ii) it disregards all other singular values than $\sigma_{\min}$ and $\sigma_{\max}$.

3.2.2 Distance from Favorable Propagation

As discussed above, when $\{\boldsymbol{g}_k\}$ have different norms or $\{\beta_k\}$ are different, we cannot use the condition number to measure how favorable the propagation is. For this case, we propose to use the *distance from favorable propagation* which is defined as the relative gap between the capacity C obtained by this propagation and the upper bound in (6):

$$\Delta_C \triangleq \frac{\sum_{k=1}^{K}\log_2\left(1+\rho\|\boldsymbol{g}_k\|^2\right) - \log_2\left|\boldsymbol{I}+\rho\boldsymbol{G}^H\boldsymbol{G}\right|}{\log_2\left|\boldsymbol{I}+\rho\boldsymbol{G}^H\boldsymbol{G}\right|}. \tag{11}$$

The distance from favorable propagation represents how far from favorable propagation the channel is. Of course, when $\Delta_C = 0$, from (6), we have favorable propagation.

4 Favorable Propagation: Rayleigh Fading and Line-of-Sight Channels

One of the key properties of Massive MIMO systems is that the channel under some conditions can offer asymptotically favorable propagation. The basic question is, under what conditions is the channel favorable? A more general question is

what practical scenarios result in favorable propagation. In practice, the channel properties depend a lot on the propagation environment as well as the antenna configurations. Therefore, there are varieties of channel models such as Rayleigh fading, Rician, finite dimensional channels, keyhole channels, LoS. In this section, we will consider two particular channel models: independent Rayleigh fading and uniform random line-of-sight (UR-LoS). These channels represent very different physical scenarios. We will study how favorable these channels are and compare the singular value spread. For simplicity, we set $\beta_k = 1$ for all k in this section.

4.1 Independent Rayleigh Fading

Consider the channel model (2) where $\{h_k^m\}$ are i.i.d. $CN(0,1)$ RVs. Note that, under a wide range of conditions, independent Rayleigh model matches the behavior of experimental data [6]. By using the law of large numbers, we have

$$\frac{1}{M}\|\boldsymbol{g}_k\|^2 \to 1, \qquad M \to \infty, \qquad \text{and} \tag{12}$$

$$\frac{1}{M}\boldsymbol{g}_k^H \boldsymbol{g}_j \to 0, \qquad M \to \infty, \qquad k \neq j, \tag{13}$$

so we have asymptotically favorable propagation.

In practice, M is large but finite. Equations (12)–(13) show the asymptotic results when $M \to \infty$. But, they do not give an account for how close to favorable propagation the channel is when M is finite. To study this fact, we consider $\mathbb{V}\text{ar}\left(\frac{1}{M}\boldsymbol{g}_k^H \boldsymbol{g}_j\right)$. For finite M, we have

$$\mathbb{V}\text{ar}\left(\frac{1}{M}\boldsymbol{g}_k^H \boldsymbol{g}_j\right) = \frac{1}{M}. \tag{14}$$

We can see that, $\frac{1}{M}\boldsymbol{g}_k^H \boldsymbol{g}_j$ is concentrated around 0 (for $k \neq j$ or 1 (for $k = j$) with variance proportional to $1/M$.

Furthermore, in Massive MIMO, the quantity $\left|\boldsymbol{g}_k^H \boldsymbol{g}_j\right|^2$ is of particular interest. For example, with matched filtering, the power of the desired signal is proportional to $\|\boldsymbol{g}_k\|^4$, while the power of the interference is proportional to $\left|\boldsymbol{g}_k^H \boldsymbol{g}_j\right|^2$, where $k \neq j$. For $k \neq j$, we have that

$$\frac{1}{M^2}|\boldsymbol{g}_k^H \boldsymbol{g}_j|^2 \to 0, \tag{15}$$

$$\mathbb{V}\text{ar}\left(\frac{1}{M^2}|\boldsymbol{g}_k^H \boldsymbol{g}_j|^2\right) = \frac{M+2}{M^3} \approx \frac{1}{M^2}. \tag{16}$$

Equation (15) shows the convergence of the random quantities $\{\left|\boldsymbol{g}_k^H \boldsymbol{g}_j\right|^2\}$ when $M \to \infty$ which represents the asymptotical favorable propagation of the channel, and (16) shows the speed of the convergence.

4.2 Uniform Random Line-of-Sight

We consider a scenario with only free space non-fading line of sight propagation between the BS and the terminals. We assume that the antenna array is uniform and linear with antenna spacing d. Then in the far-field regime, the channel vector $\boldsymbol{g}_k$ can be modelled as:

$$\boldsymbol{g}_k = e^{i\phi_k} \begin{bmatrix} 1 & e^{-i2\pi \frac{d}{\lambda} \sin(\theta_k)} & \cdots & e^{-i2\pi(M-1)\frac{d}{\lambda}\sin(\theta_k)} \end{bmatrix}^T, \tag{17}$$

where ϕ_k is uniformly distributed in $[0, 2\pi]$, θ_k is the arrival angle from the kth terminal measured relative to the array boresight, and λ is the carrier wavelength.

For any fixed and distinct angles $\{\theta_k\}$, it is straightforward to show that

$$\frac{1}{M}\|\boldsymbol{g}_k\|^2 = 1, \text{ and } \frac{1}{M}\boldsymbol{g}_k^H \boldsymbol{g}_j \to 0, \ M \to \infty, \ k \neq j, \tag{18}$$

so we have asymptotically favorable propagation.

Now assume that the K angles $\{\theta_k\}$ are randomly and independently chosen such that $\sin(\theta_k)$ is uniformly distributed in $[-1, 1]$.[1] We refer to this case as *uniformly random line-of-sight*. In this case, and if additionally $d = \lambda/2$, then

$$\mathbb{V}\text{ar}\left(\frac{1}{M}\boldsymbol{g}_k^H \boldsymbol{g}_j\right) = \frac{1}{M}. \tag{19}$$

Comparing (14) and (19), we see that the inner products between different channel vectors $\boldsymbol{g}_k$ and $\boldsymbol{g}_j$ converge to zero at the same rate for both i.i.d. Rayleigh fading and UR-LoS.

Now consider the quantity $\left|\boldsymbol{g}_k^H \boldsymbol{g}_j\right|^2$. For the UR-LoS scenario, with $k \neq j$, we have

$$\frac{1}{M^2}|\boldsymbol{g}_k^H \boldsymbol{g}_j|^2 \to 0, \tag{20}$$

$$\mathbb{V}\text{ar}\left(\frac{1}{M^2}|\boldsymbol{g}_k^H \boldsymbol{g}_j|^2\right) = \frac{(M-1)M(2M-1)}{3M^4} \approx \frac{2}{3M}. \tag{21}$$

We next compare (16) and (21). While the convergence of the inner products between $\boldsymbol{g}_k$ and $\boldsymbol{g}_j$ has the same rate in both i.i.d. Rayleigh fading and UR-LoS, the convergence of $\left|\boldsymbol{g}_k^H \boldsymbol{g}_j\right|^2$ is considerably slower in the UR-LoS case.

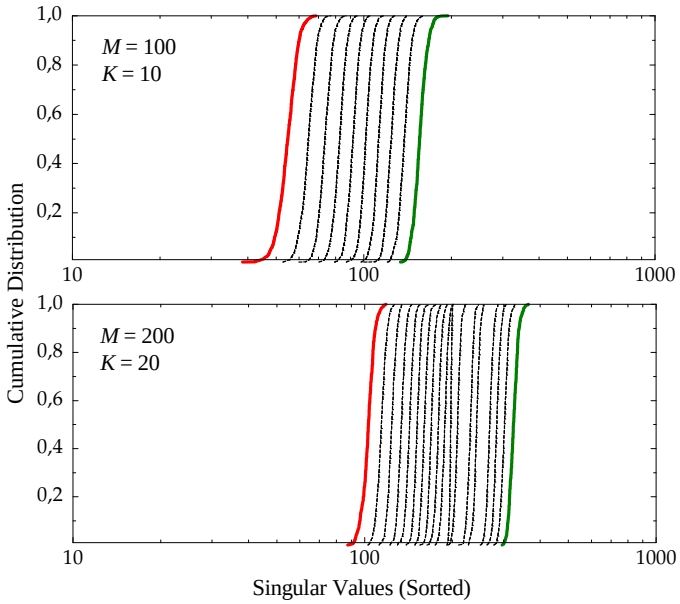


Figure 1: Singular values of $\boldsymbol{G}^H\boldsymbol{G}$ for i.i.d. Rayleigh fading. Here, $(M = 100, K = 10)$ and $(M = 200, K = 20)$.

4.3 Urns-and-Balls Model for UR-LoS

In Section 4.2, we assumed that the angles $\{\theta_k\}$ are fixed and distinct regardless of M. With this assumption, we have asymptotically favorable propagation. However, if there exist $\{\theta_k\}$ and $\{\theta_j\}$ such that $\sin(\theta_k) - \sin(\theta_j)$ is in the order of $1/M$, then we cannot have favorable propagation. To see this, assume for example that $\sin(\theta_k) - \sin(\theta_j) = 1/M$. Then

$$\frac{1}{M}\left|\boldsymbol{g}_k^H\boldsymbol{g}_j\right| = \frac{1}{M}\left|\frac{1-e^{i\pi(\sin(\theta_k)-\sin(\theta_j))M}}{1-e^{i\pi(\sin(\theta_k)-\sin(\theta_j))}}\right| = \frac{1}{M}\left|\frac{1-e^{i\pi}}{1-e^{i\pi/M}}\right| \\ \rightarrow \frac{2}{\pi} \neq 0, \quad M \rightarrow \infty. \tag{22}$$

In practice, M is finite. If the number of terminals K is in order of tens, then the probability that there exist $\{\theta_k\}$ and $\{\theta_j\}$ such that $\sin(\theta_k) - \sin(\theta_j) \leq 1/M$ cannot be neglected. This makes the channel unfavorable. This insight can be confirmed

[1] A more practical assumption is θ_k is uniformly distributed in $[0, 2\pi]$. However, it is difficult to perform analysis under this assumption, since some expressions take on an intractable form. More importantly, antennas (such as half-wavelength antenna spacing) have a directional response that discriminates against large angles of arrival, e.g., the regime where the two models ($\sin(\theta_k)$ uniformly distributed and θ_k uniformly distributed) are most different. Thus, there may be no significant difference between these two models.

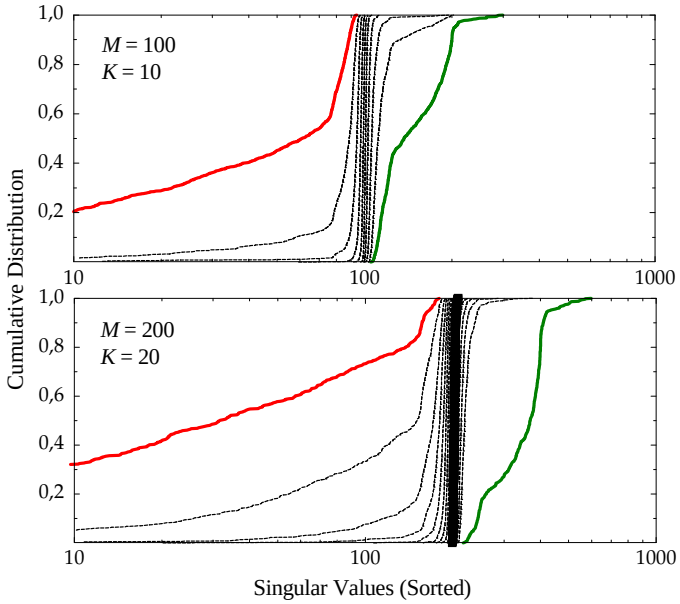


Figure 2: Same as Figure 1, but for UR-LoS.

by the following examples. Let consider the singular values of the Gramian matrix $\boldsymbol{G}^H\boldsymbol{G}$. Figures 1 and 2 show the cumulative distributions of the singular values of $\boldsymbol{G}^H\boldsymbol{G}$ for i.i.d. Raleigh fading and UR-LoS channels, respectively. We can see that in i.i.d. Rayleigh fading, the singular values are uniformly spread out between $\sigma_{\min}$ and $\sigma_{\max}$. However, for UR-LoS, two (for the case of $M = 100, K = 10$) or three (for the case of $M = 200, K = 20$) of the singular values are very small with a high probability. However, the rest are highly concentrated around their median. In order to have favorable propagation, we must drop some terminals from service.

To quantify approximately how many terminals that have to be dropped from service so that we have favorable propagation with high probability in the UR-LoS case, we propose to use the following simplified model. The BS array can create M orthogonal beams with the angles $\{\theta_m\}$:

$$\sin(\theta_m) = -1 + \frac{2m-1}{M}, \qquad m = 1, 2, ..., M. \tag{23}$$

Suppose that each one of the K terminals is randomly and independently assigned to one of the M beams given in (23). To guarantee the channel is favorable, each beam must contain at most one terminal. Therefore, if there are two or more terminals in the same beam, all but one of those terminals must be dropped from service. Let N_0, $M - K \leq N_0 < M$, be the number of beams which have no terminal. Then, the number of terminals that have to be dropped from service is

$$N_{\text{drop}} = N_0 - (M - K). \tag{24}$$

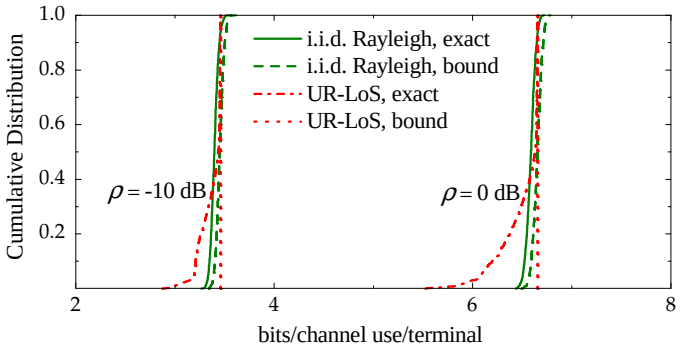


Figure 3: Capacity per terminal for i.i.d. Rayleigh fading and UR-LoS channels. Here $M = 100$ and $K = 10$.

By using a standard combinatorial result given in [7, Eq. (2.4)], we obtain the probability that n terminals, $0 \leq n < K$, are dropped as follows:

$$P\left(N_{\text{drop}} = n\right) = P\left(N_0 - (M-K) = n\right) = P\left(N_0 = n + M - K\right)$$
$$= \binom{M}{n+M-K} \sum_{k=1}^{K-n} (-1)^k \binom{K-n}{k} \left(1 - \frac{n+M-K+k}{M}\right)^K . \tag{25}$$

Therefore, the average number of terminals that must be dropped from service is

$$\bar{N}_{\text{drop}} = \sum_{n=1}^{K-1} nP\left(N_{\text{drop}} = n\right). \tag{26}$$

Remark 8 *The result obtained in this subsection yields an important insight: for Rayleigh fading, terminal selection schemes will not substantially improve the performance since the singular values are uniformly spread out. By contrast, in UR-LoS, by dropping some selected terminals from service, we can improve the worst-user performance significantly.*

5 Examples and Discussions

Figure 3 shows the cumulative distribution of the capacity per terminal for i.i.d. Rayleigh fading and UR-LoS channels, when $M = 100$ and $K = 10$. The "exact" curves are obtained by using (5), and the "bound" curves are obtained by using the upper bound (6) which is the maximum sum-capacity achieved under favorable propagation. For both Rayleigh fading and UR-LoS, the sum-capacity is very close

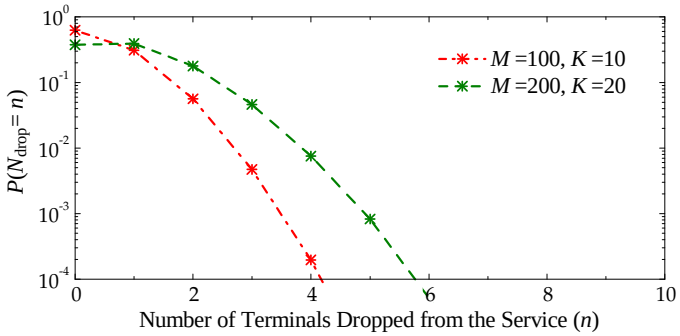


Figure 4: The probability that n terminals must be dropped from service, using the proposed urns-and-balls model.

to its upper bound with high probability. This validates our analysis: both independent Rayleigh fading and UR-LoS channels offer favorable propagation. Note that, despite the fact that the condition number for UR-LoS is large with high probability (see Fig. 1), we only need to drop a small number of terminals (2 terminals in this case) from service to have favorable propagation. As a result, the gap between capacity and its upper bound is very small with high probability.

Figure 4 shows the probability that n terminals must be dropped from service, $P\left(N_{\text{drop}} = n\right)$, for two cases: $M = 100, K = 10$ and $M = 200, K = 20$. This probability is computed by using (25). We can see that the probability that three terminals (for the case of $M = 100$, $K = 10$) and four terminals (for the case of $M = 200$, $K = 20$) must be dropped is less than 1%. This is in line with the result in Fig. 2 where three (for the case of $M = 100$, $K = 10$) or four (for the case of $M = 200$, $K = 20$) of the singular values are substantially smaller than the rest, with probability less than 1%. Note that, to guarantee favorable propagation, the number of terminals must be dropped is small ($\approx 20\%$).

6 Conclusion

Both i.i.d. Rayleigh fading and LoS with uniformly random angles-of-arrival offer asymptotically favorable propagation. In i.i.d. Rayleigh fading, the channel singular values are well spread out between the smallest and largest value. In UR-LoS, almost all singular values are concentrated around the maximum singular value, and a small number of singular values are very small. Hence, in UR-LoS, by dropping a few terminals, the propagation is approximately favorable.

The i.i.d. Rayleigh and the UR-LoS scenarios represent two extreme cases: rich scattering, and no scattering. In practice, we are likely to have a scenario which

lies in between of these two cases. Thus, it is reasonable to expect that in most practical environments, we have approximately favorable propagation.

The observations made regarding the UR-LoS model suggest that it may be worth investigating user selection schemes for massive mimo in more detail.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, Feb. 2013.

[3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, Apr. 2013.

[5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[6] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO in real propagation environments," *IEEE Trans. Wireless Commun.*, Mar. 2014, submitted.

[7] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1957, vol. 1.

Part III

System Designs

Paper D

EVD-Based Channel Estimations for Multicell Multiuser MIMO with Very Large Antenna Arrays

Refereed article Published in Proc. IEEE ICASSP 2012.

EVD-Based Channel Estimations for Multicell Multiuser MIMO with Very Large Antenna Arrays

Hien Quoc Ngo and Erik G. Larsson

Abstract

This paper consider a multicell multiuser MIMO with very large antenna arrays at the base station. For this system, with channel state information estimated from pilots, the system performance is limited by pilot contamination and noise limitation as well as the spectral inefficiency discovered in previous work. To reduce these effects, we propose the eigenvalue-decomposition-based approach to estimate the channel directly from the received data. This approach is based on the orthogonality of the channel vectors between the users and the base station when the number of base station antennas grows large. We show that the channel can be estimated from the eigenvalue of the received covariance matrix excepting the multiplicative factor ambiguity. A short training sequence is required to solved this ambiguity. Furthermore, to improve the performance of our approach, we investigate the join eigenvalue-decomposition-based approach and the Iterative Least-Square with Projection algorithm. The numerical results verify the effectiveness of our channel estimate approach.

1 Introduction

Recently, there has been a great deal of interest in multiuser MIMO (MU-MIMO) systems using very large antenna arrays (a hundred or more antennas). Such systems can provide a remarkable increase in reliability and data rate with simple signal processing [1]. When the number of base station (BS) antennas grows large, the channel vectors between the users and the BS become very long random vectors and under "favorable propagation" conditions, they become pairwisely orthogonal. As a consequence, with simple maximum-ratio combining (MRC), assuming that the BS has perfect channel state information (CSI), the interference from the other users can be cancelled without using more time-frequency resources [1]. This dramatically increases the spectral efficiency. Furthermore, by using a very large antenna array at the BS, the transmit power can be drastically reduced [2]. In [2], we showed that, with perfect CSI at the BS, we can reduce the uplink transmit power of each user inversely proportionally to the number of antennas with no reduction in performance. This holds true even with simple linear processing (MRC, or zero-forcing [ZF]) at the base station. These benefits of using large antenna arrays can be reaped if the BS has perfect CSI.

In practice, the BS does not have perfect CSI. Instead, it estimates the channels. The conventional way of doing this is to use uplink pilots. If the channel coherence time is limited, the number of possible orthogonal pilot sequences is limited too and hence, pilot sequences have to be reused in other cells. Therefore, channel estimates obtained in a given cell will be contaminated by pilots transmitted by users in other cells. This causes pilot contamination [3]. As for power efficiency, we showed in [2] that, with CSI estimated from uplink pilots, we can only reduce the uplink transmit power per user inversely proportionally to the square-root of the number of BS antennas. This is due to the fact that when we reduce the transmit power of each user, channel estimation errors will become significant. We call this effect "noise contamination". Hence, with channels estimated from pilots, the benefits of using very large antenna arrays are somewhat reduced.

In this paper we investigate whether blind channel estimation techniques could improve the performance of very large MIMO systems. Blind channel estimation techniques have been considered before as a promising approach for increasing the spectral efficiency since they require no or a minimal number of pilot symbols [4]. Generally, blind methods work well when there are unused degrees of freedom in the signal space. This is the case in very large MIMO systems, if the number of users that transmit simultaneously typically is much less than the number of antennas. One particular class of blind methods is based on a subspace partitioning of the received samples. This approach is powerful and can achieve near maximum-likelihood performance when the number of data samples is sufficiently large [5]. This approach requires a particular structure on the transmitted signal or system model, for example that the signals are coded using orthogonal space-time block codes [6, 7]. As shown later, in a system with very large antenna arrays it is possible

to apply the subspace estimation technique using eigenvalue decomposition (EVD) on the covariance matrix of the received samples, without requiring any specific structure of the transmitted signals.

The specific contributions of this paper are as follows. We consider multicell MU-MIMO systems where the BS is equipped with a very large antenna array. We propose a simple EVD-based channel estimation scheme for such systems. We show that when the number of BS antennas grows large, CSI can be estimated from the eigenvector of the covariance matrix of the received samples, up to a multiplicative scalar factor ambiguity. By using a short training sequence, this multiplicative factor ambiguity can be resolved. Finally, to improve the performance, we combine our EVD-based channel estimation technique with the iterative least-square with projection (ILSP) algorithm of [8].

2 Multi-cell Multi-user MIMO Model

Consider a multicell MU-MIMO system with L cells. Each cell contains K single-antenna users and one BS equipped with M antennas. The same frequency band is used for all L cells. We consider the uplink transmission where all users from all cells simultaneously transmit their signals to their desired BSs. Then, the $M \times 1$ received vector at the lth BS is given by[1]

$$\boldsymbol{y}_l(n) = \sqrt{p_{\mathrm{u}}} \sum_{i=1}^{L} \boldsymbol{G}_{li} \boldsymbol{x}_i(n) + \boldsymbol{n}_l(n), \tag{1}$$

where $\sqrt{p_{\mathrm{u}}}\boldsymbol{x}_i(n)$ is the $K \times 1$ vector of collectively transmitted symbols by the K users in the ith cell (the average power used by each user is p_{u}); $\boldsymbol{n}_l(n)$ is $M \times 1$ additive white noise whose elements are Gaussian with zero mean and unit variance; and $\boldsymbol{G}_{li}$ is the $M \times K$ channel matrix between the lth BS and the K users in the ith cell. The channel matrix $\boldsymbol{G}_{li}$ models independent fast fading, geometric attenuation, and log-normal shadow fading. Each element $g_{limk} \triangleq [\boldsymbol{G}_{li}]_{mk}$ is the channel coefficient between the mth antenna of the lth BS and the kth user in the ith cell, and is given by

$$g_{limk} = h_{limk}\sqrt{\beta_{lik}}, \quad m = 1, 2, ..., M, \tag{2}$$

where h_{limk} is the fast fading coefficient from the kth user in the ith cell to the mth antenna of the lth BS. We assume that h_{limk} is a random variable with zero mean and unit variance. Furthermore, $\sqrt{\beta_{lik}}$ represents the geometric attenuation and shadow fading, which are assumed to be independent of the antenna index m and to be constant and known a priori. These assumptions are reasonable since the distance between the user and the BS is much greater than the distance between

[1] When reference to n is unimportant, we will omit this index for simplicity.

the BS antennas, and the value of β_{lik} changes very slowly with time.[2] Then, the channel matrix $\boldsymbol{G}_{li}$ can be represented as

$$\boldsymbol{G}_{li} = \boldsymbol{H}_{li}\boldsymbol{D}_{li}^{1/2}, \tag{3}$$

where $\boldsymbol{H}_{li}$ is the $M \times K$ matrix of fast fading coefficients between the K users in the ith cell and the lth BS, i.e., $[\boldsymbol{H}_{li}]_{mk} = h_{limk}$, and $\boldsymbol{D}_{li}$ is a $K \times K$ diagonal matrix whose diagonal elements are $[\boldsymbol{D}_{li}]_{kk} = \beta_{lik}$.

3 EVD-based Channel Estimation

For multicell MU-MIMO systems with large antenna arrays at the BS, with conventional LS channel estimation using uplink pilots, the system performance is limited by pilot and noise contamination. Pilot contamination is caused by the interference from other cells during the training phase [1, 3]. Noise contamination occurs when the transmit power is small and the channel estimates are dominated by estimation errors [2]. Another inherent drawback of the pilot-based channel estimation is the spectral efficiency loss which results from the bandwidth consumed by training sequences. To reduce these effects, in this section, we propose an EVD-based channel estimation method.

3.1 Mathematical Preliminaries

We first consider the properties of the covariance matrix of the received vector $\boldsymbol{y}_l$. From (3) and (3), this covariance matrix is given by

$$\boldsymbol{R}_{\boldsymbol{y}} \triangleq \mathbb{E}\left\{\boldsymbol{y}_l \boldsymbol{y}_l^H\right\} = p_{\mathrm{u}} \sum_{i=1}^{L} \boldsymbol{H}_{li}\boldsymbol{D}_{li}\boldsymbol{H}_{li}^H + \boldsymbol{I}_M. \tag{4}$$

From the law of large numbers, it follows that when the number of BS antennas is large, if the fast channel coefficients are i.i.d. then the channel vectors between the users and the BS become pairwisely orthogonal, i.e.,

$$\frac{1}{M}\boldsymbol{H}_{li}^H \boldsymbol{H}_{lj} \to \delta_{ij}\boldsymbol{I}_K, \text{ as } M \to \infty. \tag{5}$$

[2] This is true assuming that the base station antennas are located, not distributed. For example, at 3 GHz, a cylindrical array comprising 4 rings of 16 dual polarized antennas elements spaced half a wavelength apart, hence having a total of 128 antennas, occupies only a physical size of 0.3×0.35 meters.

This is a key property of large MIMO systems which facilitates a simple EVD-based channel estimation that does not require any specific structure of the transmitted signals. Multiplying (4) from the right by $\boldsymbol{H}_{ll}$, and using (5), we obtain

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{y}}\boldsymbol{H}_{ll} &\approx Mp_{\mathrm{u}}\boldsymbol{H}_{ll}\boldsymbol{D}_{ll} + \boldsymbol{H}_{ll}, \text{ as } M \text{ large} \\ &= \boldsymbol{H}_{ll}\left(Mp_{\mathrm{u}}\boldsymbol{D}_{ll} + \boldsymbol{I}_K\right). \end{aligned} \tag{6}$$

For large M, the columns of $\boldsymbol{H}_{ll}$ are pairwisely orthogonal, and $Mp_{\mathrm{u}}\boldsymbol{D}_{ll} + \boldsymbol{I}_K$ is a diagonal matrix. Therefore, Equation (6) can be considered as a characteristic equation for the covariance matrix $\boldsymbol{R}_{\boldsymbol{y}}$. As a consequence, the kth column of $\boldsymbol{H}_{ll}$ is the eigenvector corresponding to the eigenvalue $Mp_{\mathrm{u}}\beta_{llk} + 1$ of $\boldsymbol{R}_{\boldsymbol{y}}$.

Remark 9 *Since $Mp_{\mathrm{u}}\beta_{llk} + 1$, $k = 1, 2, ...K$, are distinct and can be known a priori, the ordering of the eigenvectors can be determined. Each column of $\boldsymbol{H}_{ll}$ can be estimated from a corresponding eigenvector of $\boldsymbol{R}_{\boldsymbol{y}}$ up to a scalar multiplicative ambiguity.*[3] *This is due to the fact that if $\boldsymbol{u}_k$ is an eigenvector of $\boldsymbol{R}_{\boldsymbol{y}}$ corresponding to the eigenvalue $Mp_{\mathrm{u}}\beta_{llk} + 1$, then $c_k\boldsymbol{u}_k$ is also an eigenvector corresponding to that eigenvalue, for any $c_k \in \mathbb{C}$.*

Let $\boldsymbol{U}_{ll}$ be the $M \times K$ matrix whose kth column is the eigenvector of $\boldsymbol{R}_{\boldsymbol{y}}$ corresponding to the eigenvalue $Mp_{\mathrm{u}}\beta_{llk} + 1$. Then, the channel estimate of $\boldsymbol{H}_{ll}$ can be found via

$$\hat{\boldsymbol{H}}_{ll} = \boldsymbol{U}_{ll}\boldsymbol{\Xi}, \tag{7}$$

where $\boldsymbol{\Xi} \triangleq \mathrm{diag}\left\{c_1, c_2, ..., c_K\right\}$. The multiplicative matrix ambiguity $\boldsymbol{\Xi}$ can be solved by using a short pilot sequence (see Section 3.2).

3.2 Resolving the Multiplicative Factor Ambiguity

In each cell, a short training sequence of length ν symbols is used for uplink training. We assume that the training sequences of different cells are pairwisely orthogonal. Then, the $M \times \nu$ received training matrix at the lth BS is

$$\boldsymbol{Y}_{\mathrm{t},l} = \sqrt{p_{\mathrm{t}}}\boldsymbol{H}_{ll}\boldsymbol{D}_{ll}^{1/2}\boldsymbol{X}_{\mathrm{t},l} + \boldsymbol{N}_{\mathrm{t},l}, \tag{8}$$

where $\sqrt{p_{\mathrm{t}}}\boldsymbol{X}_{\mathrm{t},l}$ is the $K \times \nu$ training matrix (p_{t} is the power used by each user for each training symbol), and $\boldsymbol{N}_{\mathrm{t},l}$ is the noise matrix. From (7) and (8), the multiplicative matrix $\boldsymbol{\Xi}$ can be estimated as

$$\hat{\boldsymbol{\Xi}} = \arg\min_{\boldsymbol{\Xi}\in\boldsymbol{\Lambda}} \left\| \boldsymbol{Y}_{\mathrm{t},l} - \sqrt{p_{\mathrm{t}}}\boldsymbol{U}_{ll}\boldsymbol{\Xi}\boldsymbol{D}_{ll}^{1/2}\boldsymbol{X}_{\mathrm{t},l} \right\|_{\mathrm{F}}^2, \tag{9}$$

[3] Note that the channel matrices from other cells $\boldsymbol{H}_{li}$, for $i \neq l$, can be estimated in the same way if the large-scale fading $\boldsymbol{D}_{li}$ is known a priori.

where $\boldsymbol{\Lambda}$ is a set of $K \times K$ diagonal matrices. Denote by

$$\bar{\boldsymbol{y}}_n \triangleq \left[\left(\boldsymbol{y}_{\mathrm{t},l}^R(n)\right)^T \ \left(\boldsymbol{y}_{\mathrm{t},l}^I(n)\right)^T \right]^T,$$

where $\boldsymbol{y}_{\mathrm{t},l}(n)$ is the nth column of $\boldsymbol{Y}_{\mathrm{t},l}$, $\boldsymbol{B}^R$ and $\boldsymbol{B}^I$ denote the real and imaginary parts of matrix $\boldsymbol{B}$; and

$$\bar{\boldsymbol{A}}_n \triangleq \begin{bmatrix} \boldsymbol{A}_n^R & -\boldsymbol{A}_n^I \\ \boldsymbol{A}_n^I & \boldsymbol{A}_n^R \end{bmatrix}, \tag{10}$$

where $\boldsymbol{A}_n \triangleq \sqrt{p_{\mathrm{t}}}\boldsymbol{U}_{ll}\boldsymbol{D}_{ll}^{1/2}\bar{\boldsymbol{X}}_n$, $\bar{\boldsymbol{X}}_n \triangleq \mathrm{diag}\left(\boldsymbol{x}_{\mathrm{t},l}(n)\right)$. Then, we obtain (the proof is omitted due to space constraints)

$$\hat{\boldsymbol{\Xi}} = \mathrm{diag}\left(\hat{\boldsymbol{\xi}}\right), \tag{11}$$

where $\hat{\boldsymbol{\xi}} = [\boldsymbol{I}_K \ \ j\boldsymbol{I}_K]\hat{\bar{\boldsymbol{\xi}}}$, where

$$\hat{\bar{\boldsymbol{\xi}}} = \left(\sum_{n=1}^{\nu} \bar{\boldsymbol{A}}_n^T \bar{\boldsymbol{A}}_n\right)^{-1} \sum_{n=1}^{L} \bar{\boldsymbol{A}}_n^T \bar{\boldsymbol{y}}_n. \tag{12}$$

3.3 Implementation of the EVD-based Channel Estimation

As discussed, when M is large the channel matrix $\boldsymbol{H}_{ll}$ can be determined by using the EVD of the covariance matrix $\boldsymbol{R}_{\boldsymbol{y}}$. In practice, this covariance matrix is unavailable. Instead, we use the sample data covariance matrix $\hat{\boldsymbol{R}}_{\boldsymbol{y}}$:

$$\hat{\boldsymbol{R}}_{\boldsymbol{y}} \triangleq \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}_l(n)\,\boldsymbol{y}_l(n)^H, \tag{13}$$

where N is the number of samples. Here, we assume that the channel is still constant over at least N samples.

We summarize our proposed algorithm for estimating $\boldsymbol{H}_{ll}$ as the following algorithm. This channel estimate is performed once for each coherence time period.

Algorithm 1 ***Proposed EVD-based channel estimation method***

1. *Using a data block of N samples, compute $\hat{\boldsymbol{R}}_{\boldsymbol{y}}$ as* (13).
2. *Perform the EVD of $\hat{\boldsymbol{R}}_{\boldsymbol{y}}$. Then obtain an $M \times K$ matrix $\boldsymbol{U}_N$ whose kth column is the eigenvector corresponding to the eigenvalue which is closest to $Mp_{\mathrm{u}}\beta_{llk} + 1$.*[4]

[4] Since the eigenvalue is obtained from the sample data covariance matrix, the corresponding eigenvalue is only approximately equal to $Mp_{\mathrm{u}}\beta_{llk} + 1$.

3. *Compute the estimate* $\hat{\boldsymbol{\Xi}}$ *of the multiplicative matrix* $\boldsymbol{\Xi}$ *from* ν *pilot symbols using* (11).[5]

4. *The channel estimate of* $\boldsymbol{H}_{ll}$ *is determined as* $\tilde{\boldsymbol{H}}_{ll}=\boldsymbol{U}_N\hat{\boldsymbol{\Xi}}$.

Treating the above channel estimate as the true channel, we then use a linear detector (e.g., MRC, ZF) to detect the transmitted signals. Since the columns of the channel estimate $\tilde{\boldsymbol{H}}_{ll}$ are pairwisely orthogonal for large M, the performances of MRC and ZF detectors are the same [2].

Remark 10 *There are two main sources of errors in the channel estimate: (i) The covariance matrix error: this error is due to the use of the sample covariance matrix instead of the true covariance matrix. This error will decrease as the number of samples N increases (this requires that the coherence time is large); (ii) The error due to the channel vectors not being perfectly orthogonal as assumed in* (5)*. Our method exploits the asymptotic orthogonality of the channel vectors. This property is true only in the asymptotic regime, i.e, when $M \to \infty$. In practice, M is large but finite and hence, an error results.*

4 Joint EVD-based Method and ILSP Algorithm

As discussed above (see Remark 2), the EVD-based channel estimates will suffer from errors owing to a finite coherence time and a finite M. To reduce this error, in this section, we consider combining our EVD algorithm with the ILSP algorithm of [8].

Define the $K \times N$ matrix of transmitted signals from the K users in the ith cell and the $M \times N$ matrix of received signals at the lth BS respectively as

$$\boldsymbol{X}_i \triangleq [\boldsymbol{x}_i(1)\ \boldsymbol{x}_i(2)\ ...\ \boldsymbol{x}_i(N)], i = 1, 2, ..., L \tag{14}$$

$$\boldsymbol{Y}_l \triangleq [\boldsymbol{y}_l(1)\ \boldsymbol{y}_l(2)\ ...\ \boldsymbol{y}_l(N)]. \tag{15}$$

From (3), we have

$$\boldsymbol{Y}_l = \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{ll}\boldsymbol{X}_i + \sqrt{p_{\mathrm{u}}}\sum_{i\neq l}^{L}\boldsymbol{G}_{li}\boldsymbol{X}_i + \boldsymbol{N}_l, \tag{16}$$

where $\boldsymbol{N}_l \triangleq [\boldsymbol{n}_l(1)\ \boldsymbol{n}_l(2)\ ...\ \boldsymbol{n}_l(N)]$. Treating the last two terms of (16) as noise, and applying the ILSP algorithm in [8], we obtain an iterative algorithm that jointly estimates the channel and the transmitted data. The principle of operation of the

[5] When using (11) replace the true covariance matrix by the sample covariance matrix.

ILSP algorithm is as follows. Firstly, we assume that the channel $\boldsymbol{G}_{ll}$ is known, from an initial channel estimation procedure. The data are then detected via least-squares, projecting the solution onto the symbol constellation $\boldsymbol{\mathcal{X}}$ as

$$\hat{\boldsymbol{X}}_l = \arg \min_{\boldsymbol{X}_l \in \boldsymbol{\mathcal{X}}} \left\| \frac{1}{\sqrt{p_{\mathrm{u}}}} \boldsymbol{G}_{ll}^{\dagger} \boldsymbol{Y}_l - \boldsymbol{X}_l \right\|_F^2, \tag{17}$$

where the superscript $(\cdot)^{\dagger}$ denotes the pseudo-inverse. Next, the detected data $\hat{\boldsymbol{X}}_l$ are used as if they were equal to the true transmitted signal and the channel is re-estimated using least-squares,

$$\hat{\boldsymbol{G}}_{ll} = \frac{1}{\sqrt{p_{\mathrm{u}}}} \boldsymbol{Y}_l \hat{\boldsymbol{X}}_l^{\dagger}. \tag{18}$$

Equations (17) and (18), yield the ILSP algorithm for our problem. Applying the ILSP algorithm, and using the channel estimate based on EVD method discussed in Section 3 as the initial channel estimate, we obtain the joint EVD method and ILSP algorithm (EVD-ILSP).

Algorithm 2 ***The EVD-ILSP algorithm***

1. *Initialize* $\hat{\boldsymbol{G}}_{ll,0} = \tilde{\boldsymbol{H}}_{ll} \boldsymbol{D}_{ll}^{1/2}$ *(obtained by using the EVD-based method). Choose number of iterations* K_{step}. *Set* $k = 0$.

2. $k := k + 1$

- $\hat{\boldsymbol{X}}_{l,k} = \arg\min_{\boldsymbol{X}_l \in \boldsymbol{\mathcal{X}}} \left\| \frac{1}{\sqrt{p_{\mathrm{u}}}} \hat{\boldsymbol{G}}_{ll,k-1}^{\dagger} \boldsymbol{Y}_l - \boldsymbol{X}_l \right\|_F^2$
- $\hat{\boldsymbol{G}}_{ll,k} = \frac{1}{\sqrt{p_{\mathrm{u}}}} \boldsymbol{Y}_l \hat{\boldsymbol{X}}_{l,k}^{\dagger}$

3. *Repeat 2 until* $k = K_{\mathrm{step}}$.

5 Numerical Results

We simulate a system with $L = 3$ cells, each containing 3 users. We consider the SEP for the uplink of the 1st user in 1st cell, assuming BPSK modulation and ZF receivers. We choose $\boldsymbol{D}_{11} = \mathrm{diag}\{0.98, 0.63, 0.47\}$, $\boldsymbol{D}_{12} = a \times \mathrm{diag}\{0.36, 0.29, 0.05\}$, and $\boldsymbol{D}_{13} = a \times \mathrm{diag}\{0.32, 0.14, 0.11\}$. For the EVD-based method, we use $\nu = 1$ (one) training symbol to resolve the multiplicative factor ambiguity. For the pilot-based method, we perform the least-square estimation scheme using 3 symbols for pilots.

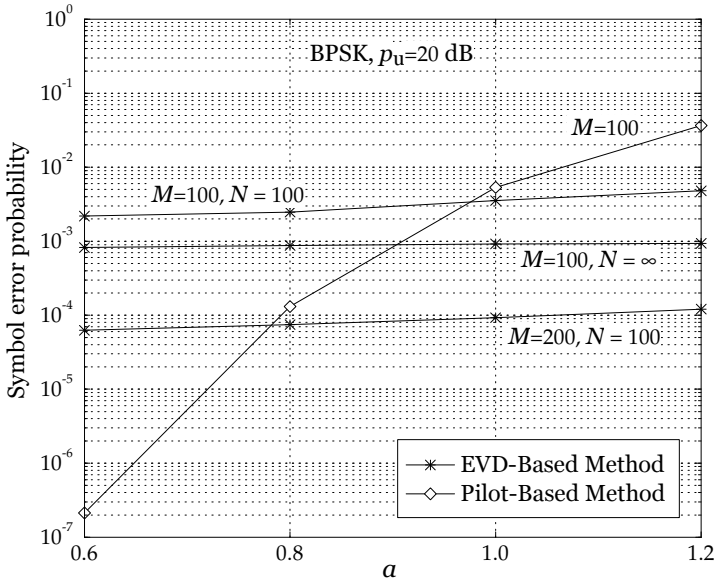


Figure 1: Symbol error probability versus a at $p_u = 20$ dB, and BPSK modulation.

Fig. 1 shows the the SEP versus a of the EVD-based and the conventional pilot-based channel estimation methods with different N and M at $p_u = 20$ dB. ($N = \infty$ implies that we have perfect knowledge of the true covariance matrix.) We can see that when a increases (the effect of pilot contamination increases), the system performance degrades dramatically when using the pilot-based method. This is due to the fact that the pilot-based method suffers from pilot contamination. In particular, the EVD-based method is not affected much by the pilot contamination, and it can significantly improve the system performance when the effect of pilot contamination is large. It can be also seen from the figure that the effectiveness of our EVD-based method increases when the number of samples N and the number of BS antennas M increase.

To ascertain the effectiveness of the EVD-based channel estimation method under noise-limited conditions, we consider the SEP when the transmit power of each user is chosen to be proportional to $1/M$. We choose $M = 100$, and $a = 1$. Fig. 2 shows the comparisons between the SEPs versus SNR of the EVD-based method and the pilot-aided method for different N. Here, with each SNR, we set $p_u = \mathsf{SNR}/M$. We can see that by using the EVD-based method, the system performance significantly improves compared with the conventional pilot-based method. When N increases, the sample covariance matrix tends to the true covariance matrix and hence, as we can see from the figure, the SEP decreases.

Fig. 3 shows the SEP of the EVD-based method versus the number of BS antennas at $p_u = 20$ dB and $a = 1$, for different N, with and without using the ILSP algorithm.

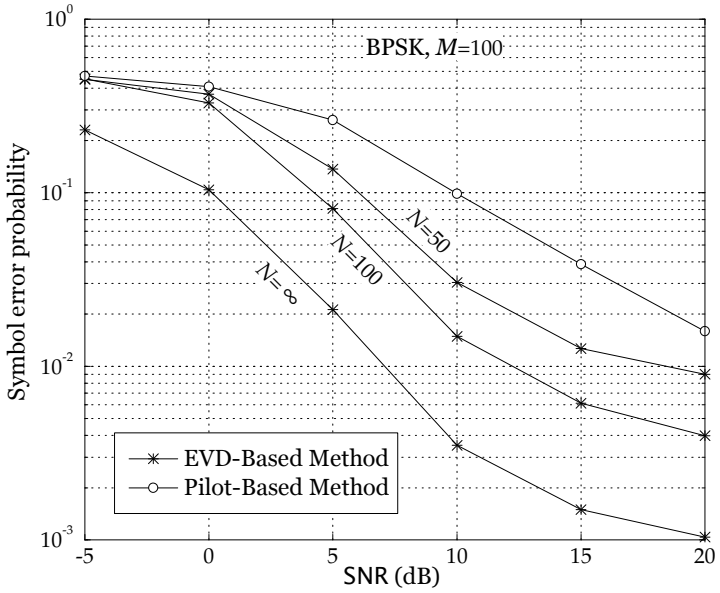


Figure 2: Symbol error probability versus SNR for $M = 100$, $a = 1$, $p_{\mathrm{u}} = \mathsf{SNR}/M$, and BPSK modulation.

With the ILSP algorithm, we choose $K_{\mathrm{step}} = 5$. As expected, comparing with the EVD-based method, the joint EVD-based and ILSP algorithm offers a performance improvement. Also here, the system performance improves significantly when M and N increase.

6 Concluding Remarks

Very large MIMO systems with $M \gg K \gg 1$ offer many unused degrees of freedom. We proposed a channel estimation method that exploits these excess degrees of freedom, together with the asymptotic orthogonality between the channel vectors that occurs under "favorable propagation" conditions. Combining the proposed method with the ILSP algorithm of [8] further enhances performance.

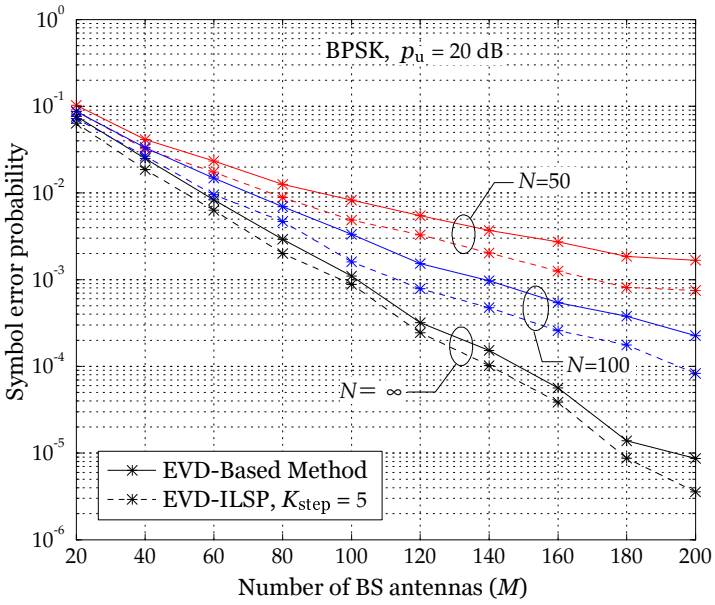


Figure 3: Symbol error probability versus M for $p_u = 20$ dB, and $a = 1$.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Proc. 49th Allerton Conference on Communication, Control, and Computing*, 2011.

[3] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[4] A.-J. van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Process. Lett.*, vol. 2, no. 5, pp. 99–102, May 1995.

[5] A.-J. van der Veen, S. Talwar, and A. Paulraj, "A subsapce approach to blind space-time signal processing for wireless communication systems," *IEEE Signal Process. Lett.*, vol. 45, no. 1, pp. 173–189, Jan. 1997.

[6] E. Beres and R. Adve, "Blind channel estimation for orthogonal STBC in MISO systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2042–2050, 2007.

[7] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, 2002.

[8] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—part I: Algorithms," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1184–1197, May 1996.

Paper E

Massive MU-MIMO Downlink TDD Systems with Linear Precoding and Downlink Pilots

Refereed article Published in Proc. IEEE ACCCC 2013.

Massive MU-MIMO Downlink TDD Systems with Linear Precoding and Downlink Pilots

Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta

Abstract

We consider a massive MU-MIMO downlink time-division duplex system where a base station (BS) equipped with many antennas serves several single-antenna users in the same time-frequency resource. We assume that the BS uses linear precoding for the transmission. To reliably decode the signals transmitted from the BS, each user should have an estimate of its channel. In this work, we consider an efficient channel estimation scheme to acquire CSI at each user, called beamforming training *scheme. With the beamforming training scheme, the BS precodes the pilot sequences and forwards to all users. Then, based on the received pilots, each user uses minimum mean-square error channel estimation to estimate the effective channel gains. The channel estimation overhead of this scheme does not depend on the number of BS antennas, and is only proportional to the number of users. We then derive a lower bound on the capacity for maximum-ratio transmission and zero-forcing precoding techniques which enables us to evaluate the spectral efficiency taking into account the spectral efficiency loss associated with the transmission of the downlink pilots. Comparing with previous work where each user uses only the statistical channel properties to decode the transmitted signals, we see that the proposed beamforming training scheme is preferable for moderate and low-mobility environments.*

1 Introduction

Recently, massive (or very large) multiuser multiple-input multiple-output (MU-MIMO) systems have attracted a lot of attention from both academia and industry [1–4]. Massive MU-MIMO is a system where a base station (BS) equipped with many antennas simultaneously serves several users in the same frequency band. Owing to the large number of degrees-of-freedom available for each user, massive MU-MIMO can provide a very high data rate and communication reliability with simple linear processing such as maximum-ratio combining (MRC) or zero-forcing (ZF) on the uplink and maximum-ratio transmission (MRT) or ZF on the downlink. At the same time, the radiated energy efficiency can be significantly improved [5]. Therefore, massive MU-MIMO is considered as a promising technology for next generations of cellular systems. In order to use the advantages that massive MU-MIMO can offer, accurate channel state information (CSI) is required at the BS and/or the users.

In small MU-MIMO systems where the number of BS antennas is relatively small, typically, the BS can acquire an estimate of CSI via feedback in frequency-division duplex (FDD) operation [6]. More precisely, each user estimates the channels based on the downlink training, and then it feeds back its channel estimates to the BS through the reverse link. However, in massive MU-MIMO systems, the number of BS antennas is very large and channel estimation becomes challenging in FDD since the number of downlink resources needed for pilots will be proportional to the number of BS antennas. Also, the required bandwidth for CSI feedback becomes very large. By contrast, in time-division duplex (TDD) systems, owing to the channel reciprocity, the BS can obtain CSI in open-loop directly from the uplink training. The pilot transmission overhead is thus proportional to the number of users which is typically much smaller than the number of BS antennas. Therefore, CSI acquisition at the BS via open-loop training under TDD operation is preferable in massive MU-MIMO systems [1–3, 7, 8]. With this CSI acquisition, in the uplink, the signals transmitted from the users can be decoded by using these channel estimates. In the downlink, the BS can use the channel estimates to precode the transmit signals. However, the channel estimates are only available at the BS. The user also should have an estimate of the channel in order to reliably decode the transmitted signals in the downlink. To acquire CSI at the users, a simple scheme is that the BS sends the pilots to the users. Then, each user will estimate the channel based on the received pilots. This is very inefficient since the channel estimation overhead will be proportional to the number of BS antennas. Therefore, the majority of the research on these systems has assumed that the users do not have knowledge of the CSI. More precisely, the signal is detected at each user by only using the statistical properties of the channels [7–9]. Some work assumed that the users have perfect CSI [10]. To the authors' best knowledge, it has not been previously considered how to efficiently acquire CSI at each user in the massive MU-MIMO downlink.

In this paper, we propose a beamforming training scheme to acquire estimates of the CSI at each user. With this scheme, instead of forwarding a long pilot sequence

(whose length is proportional to the number of BS antennas), the BS just beamforms a short pilot sequence so that each user can estimate the effective channel gain (the combination of the precoding vector and the channel gain). The channel estimation overhead of this scheme is only proportional to the number of users. To evaluate the performance of the proposed beamforming training scheme, we derive a lower bound on the capacity of two specific linear precoding techniques, namely MRT and ZF. Numerical results show that the beamforming training scheme works very well in moderate and low-mobility environments.

Notation: We use upper (lower) bold letters to denote matrices (vectors). The superscripts T, $*$, and H stand for the transpose, conjugate, and conjugate-transpose, respectively. $\mathrm{tr}\,(\boldsymbol{A})$ denotes the trace of a matrix $\boldsymbol{A}$, and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. The expectation operator and the Euclidean norm are denoted by $\mathbb{E}\{\cdot\}$ and $\|\cdot\|$, respectively. Finally, we use $\boldsymbol{z} \sim \mathcal{CN}(0, \boldsymbol{\Sigma})$ to denote a circularly symmetric complex Gaussian vector $\boldsymbol{z}$ with zero mean and covariance matrix $\boldsymbol{\Sigma}$.

2 System Model and Beamforming Training

We consider the downlink transmission in a MU-MIMO system where a BS equipped with M antennas serves K single-antenna users in the same time-frequency resource, see Fig. 1. Here, we assume that $M \gg K$. We further assume that the BS uses linear precoding techniques to process the signal before transmitting to all users. This requires knowledge of CSI at the BS. We assume TDD operation so that the channels on the uplink and downlink are equal. The estimates of CSI are obtained from uplink training.

2.1 Uplink Training

Let τ_{u} be the number of symbols per coherence interval used entirely for uplink pilots. All users simultaneously transmit pilot sequences of length τ_{u} symbols. The pilot sequences of K users are pairwisely orthogonal. Therefore, it is required that $\tau_{\mathrm{u}} \geq K$.

Denote by $\boldsymbol{H} \in \mathbb{C}^{M \times K}$ the channel matrix between the BS and the K users. We assume that elements of $\boldsymbol{H}$ are i.i.d. Gaussian distributed with zero mean and unit variance. Here, for the simplicity, we neglect the effects of large-scale fading. Then, the minimum mean-square error (MMSE) estimate of $\boldsymbol{H}$ is given by [11]

$$\hat{\boldsymbol{H}} = \frac{\tau_{\mathrm{u}} p_{\mathrm{u}}}{\tau_{\mathrm{u}} p_{\mathrm{u}} + 1} \boldsymbol{H} + \frac{\sqrt{\tau_{\mathrm{u}} p_{\mathrm{u}}}}{\tau_{\mathrm{u}} p_{\mathrm{u}} + 1} \boldsymbol{N}_{\mathrm{u}}, \tag{1}$$

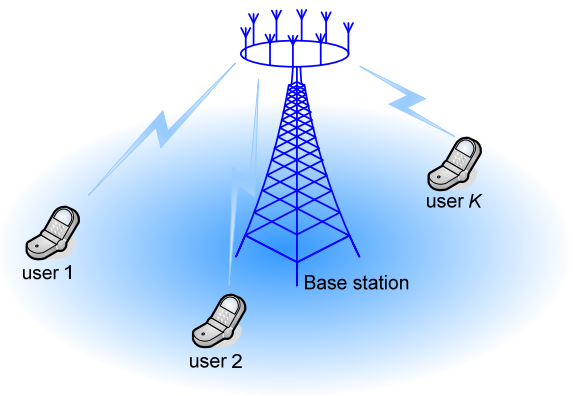


Figure 1: Massive MU-MIMO downlink system model.

where $\boldsymbol{N}_{\mathrm{u}}$ is a Gaussian matrix with i.i.d. $\mathcal{CN}(0,1)$ entries, and p_{u} denotes the average transmit power of each uplink pilot symbol. The channel matrix $\boldsymbol{H}$ can be decomposed as

$$\boldsymbol{H} = \hat{\boldsymbol{H}} + \boldsymbol{\mathcal{E}}, \tag{2}$$

where $\boldsymbol{\mathcal{E}}$ is the channel estimation error. Since we use MMSE channel estimation, $\hat{\boldsymbol{H}}$ and $\boldsymbol{\mathcal{E}}$ are independent [11]. Furthermore, $\hat{\boldsymbol{H}}$ has i.i.d. $\mathcal{CN}\left(0, \frac{\tau_{\mathrm{u}} p_{\mathrm{u}}}{\tau_{\mathrm{u}} p_{\mathrm{u}}+1}\right)$ elements, and $\boldsymbol{\mathcal{E}}$ has i.i.d. $\mathcal{CN}\left(0, \frac{1}{\tau_{\mathrm{u}} p_{\mathrm{u}}+1}\right)$ elements.

2.2 Downlink Transmission

Let s_k be the symbol to be transmitted to the kth user, with $\mathbb{E}\left\{|s_k|^2\right\} = 1$. The BS uses the channel estimate $\hat{\boldsymbol{H}}$ to linearly precode the symbols, and then it transmits the precoded signal vector to all users. Let $\boldsymbol{W} \in \mathbb{C}^{M \times K}$ be the linear precoding matrix which is a function of the channel estimate $\hat{\boldsymbol{H}}$. Then, the $M \times 1$ transmit signal vector is given by

$$\boldsymbol{x} = \sqrt{p_{\mathrm{d}}} \boldsymbol{W} \boldsymbol{s}, \tag{3}$$

where $\boldsymbol{s} \triangleq [s_1 \ s_2 \ \ldots \ s_K]^T$, and p_{d} is the average transmit power at the BS. To satisfy the power constraint at the BS, $\boldsymbol{W}$ is chosen such as $\mathbb{E}\left\{\|\boldsymbol{x}\|^2\right\} = p_{\mathrm{d}}$, or equivalently $\mathbb{E}\left\{\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{W}^H\right)\right\} = 1$.

The vector of samples collectively received at the K users is given by

$$\boldsymbol{y} = \boldsymbol{H}^T \boldsymbol{x} + \boldsymbol{n} = \sqrt{p_{\mathrm{d}}} \boldsymbol{H}^T \boldsymbol{W} \boldsymbol{s} + \boldsymbol{n}, \tag{4}$$

where $\boldsymbol{n}$ is a vector whose kth element, n_k, is the additive noise at the kth user. We assume that $n_k \sim \mathcal{CN}(0,1)$. Define $a_{ki} \triangleq \boldsymbol{h}_k^T \boldsymbol{w}_i$, where $\boldsymbol{h}_i$ and $\boldsymbol{w}_i$ are the ith columns of $\boldsymbol{H}$ and $\boldsymbol{W}$, respectively. Then, the received signal at the kth user can be written as

$$y_k = \sqrt{p_{\mathrm{d}}} a_{kk} s_k + \sqrt{p_{\mathrm{d}}} \sum_{i \neq k}^{K} a_{ki} s_i + n_k. \tag{5}$$

Remark 11 *Each user should have CSI to coherently detect the transmitted signals. A simple way to acquire CSI is to use downlink pilots. The channel estimate overhead will be proportional to M. In massive MIMO, M is large, so it is inefficient to estimate the full channel matrix $\boldsymbol{H}$ at each user using downlink pilots. This is the reason for why most of previous studies assumed that the users have only knowledge of the statistical properties of the channels [8,9]. More precisely, in [8,9], the authors use $\mathbb{E}\{a_{kk}\}$ to detect the transmitted signals. With very large M, a_{kk} becomes nearly deterministic. In this case, using $\mathbb{E}\{a_{kk}\}$ for the signal detection is good enough. However, for moderately large M, the users should have CSI in order to reliably decode the transmitted signals. We observe from* (5) *that to detect s_k, user k does not need the knowledge of $\boldsymbol{H}$ (which has a dimension of $M \times K$). Instead, user k needs only to know a_{kk} which is a scalar value. Therefore, to acquire a_{kk} at each user, we can spend a small amount of the coherence interval on downlink training. In the next section, we will provide more detail about this proposed downlink* beamforming training *scheme to estimate a_{kk}. With this scheme, the channel estimation overhead is proportional to the number of users K.*

2.3 Beamforming Training Scheme

The BS beamforms the pilots. Then, the kth user will estimate a_{ki} by using the received pilots. Let $\boldsymbol{S}_{\mathrm{p}} \in \mathbb{C}^{K \times \tau_d}$ be the pilot matrix, where τ_{d} is the duration (in symbols) of the downlink training. The pilot matrix is given by

$$\boldsymbol{S}_{\mathrm{p}} = \sqrt{\tau_{\mathrm{d}} p_{\mathrm{d}}} \boldsymbol{\Phi}. \tag{6}$$

We assume that the rows of $\boldsymbol{\Phi}$ are pairwisely orthonormal, i.e., $\boldsymbol{\Phi}\boldsymbol{\Phi}^H = \boldsymbol{I}_K$. This requires that $\tau_d \geq K$.

The BS beamforms the pilot sequence using the precoding matrix $\boldsymbol{W}$. More precisely, the transmitted pilot matrix is $\boldsymbol{W}\boldsymbol{S}_{\mathrm{p}}$. Then, the $K \times \tau_{\mathrm{d}}$ received pilot matrix at the K users is given by

$$\boldsymbol{Y}_{\mathrm{p}}^T = \sqrt{\tau_{\mathrm{d}} p_{\mathrm{d}}} \boldsymbol{H}^T \boldsymbol{W} \boldsymbol{\Phi} + \boldsymbol{N}_{\mathrm{p}}^T. \tag{7}$$

where $\boldsymbol{N}_{\mathrm{p}}$ is the AWGN matrix whose elements are i.i.d. $\mathcal{CN}(0,1)$. The received pilot matrix $\boldsymbol{Y}_{\mathrm{p}}^T$ can be represented by $\boldsymbol{Y}_{\mathrm{p}}^T \boldsymbol{\Phi}^H$ and $\boldsymbol{Y}_{\mathrm{p}}^T \boldsymbol{\Phi}_{\perp}^H$, where $\boldsymbol{\Phi}_{\perp}^H$ is the orthogonal complement of $\boldsymbol{\Phi}^H$, i.e., $\boldsymbol{\Phi}_{\perp}^H = \boldsymbol{I}_{\tau_{\mathrm{d}}} - \boldsymbol{\Phi}^H \boldsymbol{\Phi}$. We can see that $\boldsymbol{Y}_{\mathrm{p}}^T \boldsymbol{\Phi}_{\perp}^H$ only

includes noise which is independent of $\boldsymbol{Y}_{\mathrm{p}}^T\boldsymbol{\Phi}^H$. Thus, it is sufficient to use $\boldsymbol{Y}_{\mathrm{p}}^T\boldsymbol{\Phi}^H$ for the channel estimation. Let

$$\tilde{\boldsymbol{Y}}_{\mathrm{p}}^T \triangleq \boldsymbol{Y}_{\mathrm{p}}^T\boldsymbol{\Phi}^H = \sqrt{\tau_{\mathrm{d}}p_{\mathrm{d}}}\boldsymbol{H}^T\boldsymbol{W} + \tilde{\boldsymbol{N}}_{\mathrm{p}}^T, \tag{8}$$

where $\tilde{\boldsymbol{N}}_{\mathrm{p}}^T \triangleq \boldsymbol{N}_{\mathrm{p}}^T\boldsymbol{\Phi}^H$ has i.i.d. $\mathcal{CN}(0,1)$ elements. From (8), the $1 \times K$ received pilot vector at user k is given by

$$\tilde{\boldsymbol{y}}_{\mathrm{p},k}^T = \sqrt{\tau_{\mathrm{d}}p_{\mathrm{d}}}\boldsymbol{h}_k^T\boldsymbol{W} + \tilde{\boldsymbol{n}}_{\mathrm{p},k}^T = \sqrt{\tau_{\mathrm{d}}p_{\mathrm{d}}}\boldsymbol{a}_k^T + \tilde{\boldsymbol{n}}_{\mathrm{p},k}^T, \tag{9}$$

where $\boldsymbol{a}_k \triangleq [a_{k1}\ a_{k2}\ ...\ a_{kK}]^T$, and $\tilde{\boldsymbol{y}}_{\mathrm{p},k}$ and $\tilde{\boldsymbol{n}}_{\mathrm{p},k}$ are the kth columns of $\tilde{\boldsymbol{Y}}_{\mathrm{p}}$ and $\tilde{\boldsymbol{N}}_{\mathrm{p}}$, respectively.

From the received pilot $\tilde{\boldsymbol{y}}_{\mathrm{p},k}^T$, user k estimates $\boldsymbol{a}_k$. Depending on the precoding matrix $\boldsymbol{W}$, the elements of $\boldsymbol{a}_k$ can be correlated and hence, they should be jointly estimated. However, here, for the simplicity of the analysis, we estimate $a_{k1}, ..., a_{kK}$ independently, i.e., we use the ith element of $\tilde{\boldsymbol{y}}_{p,k}$ to estimate a_{ki}. In Section 4, we show that estimating the elements of $\boldsymbol{a}_k$ jointly will not improve the system performance much compared to the case where the elements of $\boldsymbol{a}_k$ are estimated independently. The MMSE channel estimate of a_{ki} is given by [11]

$$\hat{a}_{ki} = \mathbb{E}\{a_{ki}\} + \frac{\sqrt{\tau_{\mathrm{d}}p_{\mathrm{d}}}\mathtt{Var}(a_{ki})}{\tau_{\mathrm{d}}p_{\mathrm{d}}\mathtt{Var}(a_{ki}) + 1}\left(\tilde{y}_{\mathrm{p},ki} - \sqrt{\tau_{\mathrm{d}}p_{\mathrm{d}}}\,\mathbb{E}\{a_{ki}\}\right), \tag{10}$$

where $\mathtt{Var}(a_{ki}) \triangleq \mathbb{E}\left\{|a_{ki} - \mathbb{E}\{a_{ki}\}|^2\right\}$, and $\tilde{y}_{\mathrm{p},ki}$ is the ith element of $\tilde{\boldsymbol{y}}_{\mathrm{p},k}$. Let ϵ_{ki} be the channel estimation error. Then, the effective channel a_{ki} can be decomposed as

$$a_{ki} = \hat{a}_{ki} + \epsilon_{ki}. \tag{11}$$

Note that, since we use MMSE estimation, the estimate $\hat{a}_{ki}$ and the estimation error ϵ_{ki} are uncorrelated.

3 Achievable Downlink Rate

In this section, we derive a lower bound on the achievable downlink rate for MRT and ZF precoding techniques, using the proposed beamforming training scheme. To obtain these achievable rates, we use the techniques of [12].

User k uses the channel estimate $\hat{\boldsymbol{a}}_k$ in (10) to detect the transmitted signal s_k. Therefore, the achievable downlink rate of the transmission from the BS to the kth user is the mutual information between the unknown transmitted signal s_k and the observed received signal y_k given by (5) and the known channel estimate $\hat{\boldsymbol{a}}_k = [\hat{a}_{k1}\ ...\ \hat{a}_{kK}]^T$, i.e., $I(s_k; y_k, \hat{\boldsymbol{a}}_k)$.

Following a similar methodology as in [12, Appendix A], we obtain a lower bound on the achievable rate of the transmission from the BS to the kth user as:

$$R_k = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{d}} \left|\hat{a}_{kk}\right|^2}{p_{\mathrm{d}} \sum_{i=1}^{K} \mathbb{E}\left\{\left|\epsilon_{ki}\right|^2\right\} + p_{\mathrm{d}} \sum_{i \neq k}^{K} \left|\hat{a}_{ki}\right|^2 + 1}\right)\right\}. \tag{12}$$

We next simplify the capacity lower bound given by (12) for two specific linear precoding techniques at the BS, namely, MRT and ZF.

3.1 Maximum-Ratio Transmission

With MRT, the precoding matrix $\boldsymbol{W}$ is given by

$$\boldsymbol{W} = \alpha_{\mathtt{MRT}} \hat{\boldsymbol{H}}^*, \tag{13}$$

where $\alpha_{\mathtt{MRT}}$ is a normalization constant chosen to satisfy the transmit power constraint at the BS, i.e., $\mathbb{E}\left\{\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{W}^H\right)\right\} = 1$. Hence,

$$\alpha_{\mathtt{MRT}} = \sqrt{\frac{1}{\mathbb{E}\left\{\mathrm{tr}\left(\hat{\boldsymbol{H}}^* \hat{\boldsymbol{H}}^T\right)\right\}}} = \sqrt{\frac{\tau_{\mathrm{u}} p_{\mathrm{u}} + 1}{MK\tau_{\mathrm{u}} p_{\mathrm{u}}}}. \tag{14}$$

Proposition 10 *With MRT, the lower bound on the achievable rate given by* (12) *becomes*

$$R_k = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{d}} \left|\hat{a}_{kk}\right|^2}{\frac{Kp_{\mathrm{d}}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K} + p_{\mathrm{d}} \sum_{i \neq k}^{K} \left|\hat{a}_{ki}\right|^2 + 1}\right)\right\}, \tag{15}$$

where

$$\hat{a}_{ki} = \frac{\sqrt{\tau_{\mathrm{d}} p_{\mathrm{d}}}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K} \tilde{\boldsymbol{y}}_{p,ki} + \frac{K}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K} \sqrt{\frac{\tau_{\mathrm{u}} p_{\mathrm{u}} M}{K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}} \delta_{ki}, \tag{16}$$

where $\delta_{ki} = 1$ *when* $i = k$ *and* 0 *otherwise.*

Proof: *See Appendix A.* □

3.2 Zero-Forcing

With ZF, the precoding matrix is

$$\boldsymbol{W} = \alpha_{\mathtt{ZF}} \hat{\boldsymbol{H}}^* \left(\hat{\boldsymbol{H}}^T \hat{\boldsymbol{H}}^* \right)^{-1}, \tag{17}$$

where the normalization constant $\alpha_{\mathtt{ZF}}$ is chosen to satisfy the power constraint $\mathbb{E}\left\{ \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{W}^H \right) \right\} = 1$, i.e., [9]

$$\alpha_{\mathtt{ZF}} = \sqrt{\frac{(M-K)\,\tau_{\mathrm{u}} p_{\mathrm{u}}}{K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}}. \tag{18}$$

Proposition 11 *With ZF, the lower bound on the achievable rate given by* (12) *becomes*

$$R_k = \mathbb{E}\left\{ \log_2 \left(1 + \frac{p_{\mathrm{d}} \left| \hat{a}_{kk} \right|^2}{\frac{K p_{\mathrm{d}}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1)} + p_{\mathrm{d}} \sum_{i \neq k}^{K} \left| \hat{a}_{ki} \right|^2 + 1} \right) \right\}, \tag{19}$$

where

$$\hat{a}_{ki} = \frac{\sqrt{\tau_{\mathrm{d}} p_{\mathrm{d}}}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)} \tilde{\boldsymbol{y}}_{p,ki} + \frac{\sqrt{K\left(M-K\right)\tau_{\mathrm{u}} p_{\mathrm{u}} \left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)} \delta_{ki}. \tag{20}$$

Proof: *See Appendix B.* □

4 Numerical Results

In this section, we illustrate the spectral efficiency performance of the beamforming training scheme. The spectral efficiency is defined as the sum-rate (in bits) per channel use. Let T be the length of the coherence interval (in symbols). During each coherence interval, we spend τ_{u} symbols for uplink training and τ_{d} symbols for beamforming training. Therefore, the spectral efficiency is given by

$$\mathcal{S}_{\mathtt{TB}} = \frac{T - \tau_{\mathrm{u}} - \tau_{\mathrm{d}}}{T} \sum_{k=1}^{K} R_k, \tag{21}$$

where R_k is given by (15) for MRT, and (19) for ZF.

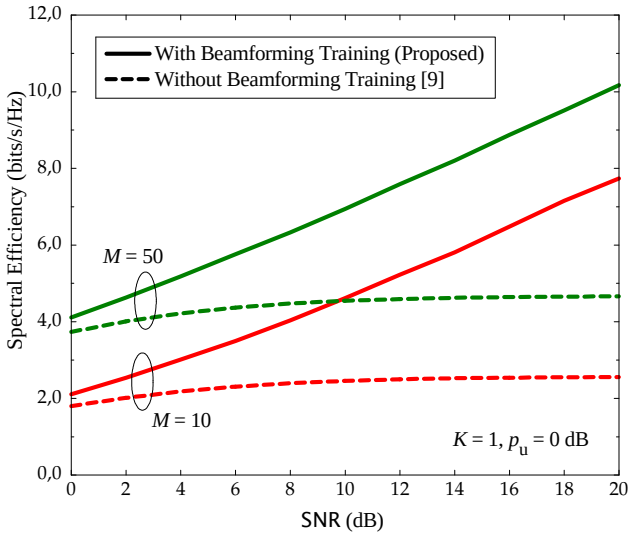


Figure 2: Spectral efficiency versus SNR for a single-user setup ($K = 1$, $p_{\mathrm{u}} = 0$ dB, and $T = 200$).

For comparison, we also consider the spectral efficiency for the case that there is no beamforming training and that $\mathbb{E}\{a_{kk}\}$ is used instead of a_{kk} for the detection [9]. The spectral efficiency for this case is given by [9]

$$\mathcal{S}_0 = \begin{cases} \frac{T-\tau_{\mathrm{u}}}{T} K \log_2\left(1 + \frac{M}{K} \frac{\tau_{\mathrm{u}} p_{\mathrm{u}} p_{\mathrm{d}}}{(p_{\mathrm{d}}+1)(\tau_{\mathrm{u}} p_{\mathrm{u}}+1)}\right), \text{for MRT} \\ \frac{T-\tau_{\mathrm{u}}}{T} K \log_2\left(1 + \frac{M-K}{K} \frac{\tau_{\mathrm{u}} p_{\mathrm{u}} p_{\mathrm{d}}}{\tau_{\mathrm{u}} p_{\mathrm{u}}+p_{\mathrm{d}}+1}\right), \text{for ZF.} \end{cases} \tag{22}$$

In all examples, we choose $\tau_{\mathrm{u}} = \tau_{\mathrm{d}} = K$ and $p_{\mathrm{u}} = 0$dB. We define $\mathsf{SNR} \triangleq p_{\mathrm{d}}$.

We first consider a single-user setup ($K = 1$). When $K = 1$, the performances MRT and ZF are the same. Fig. 1 shows the spectral efficiency versus SNR for different number of BS antennas $M = 10$ and $M = 50$, at $T = 200$ (e.g. 1ms×200kHz). We can see that the beamforming training scheme outperforms the case without beamforming training. The performance gap increases significantly when the SNR increases. The reason is that, when SNR (or the downlink power) increases, the channel estimate at each user is more accurate and hence, the advantage of the beamforming training scheme grows.

Next, we consider a multiuser setup. Here, we choose the number of users to be $K = 5$. Fig. 2 shows the spectral efficiency versus SNR for the MRT and ZF precoders, at $M = 10$, $M = 50$, and $T = 200$. Again, the beamforming training offers an improved performance. In addition, we can see that the beamforming training with MRT precoding is more efficient than the beamforming training with ZF precoding. This is due to the fact that, with ZF, the randomness of the effective

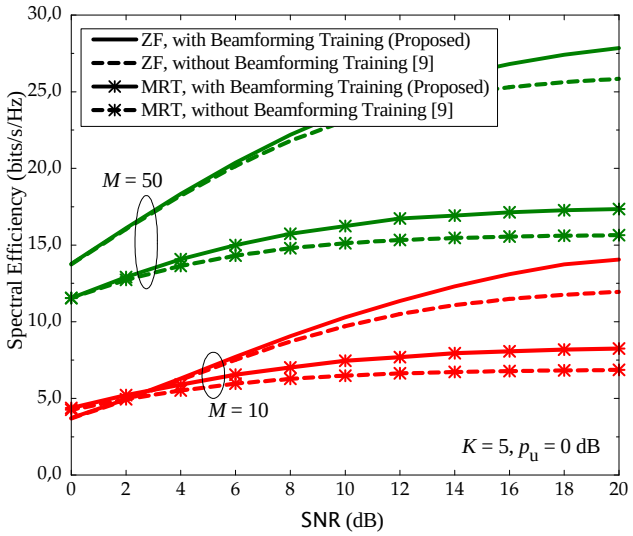


Figure 3: Spectral efficiency versus SNR for a multiuser setup ($K = 5$, $p_{\mathrm{u}} = 0$ dB, and $T = 200$).

channel gain a_{kk} at the kth user is smaller than the one with MRT (with ZF, a_{kk} becomes deterministic when the BS has perfect CSI) and hence, MRC has a higher advantage of using the channel estimate for the signal detection.

Furthermore, we consider the effect of the length of the coherence interval on the system performance of the beamforming training scheme. Fig. 3 shows the spectral efficiency versus the length of the coherence interval T at $M = 50$, $K = 5$, and $p_{\mathrm{d}} = 20$ dB. As expected, for short coherence intervals (in a high-mobility environment), the training duration is relatively large compared to the length of the coherence interval and hence, we should not use the beamforming training to estimate CSI at each user. At moderate and large T, the training duration is relatively small compared with the coherence interval. As a result, the beamforming training scheme is preferable.

Finally, we consider the spectral efficiency of our scheme but with a genie receiver, i.e., we assume that the kth user can estimate perfectly $\boldsymbol{a}_k$ in the beamforming training phase. For this case, the spectral efficiency is given by

$$\mathcal{S}_{\mathtt{G}} = \frac{T - \tau_{\mathrm{u}} - \tau_{\mathrm{d}}}{T} \sum_{k=1}^{K} \mathbb{E}\left\{ \log_2 \left(1 + \frac{p_{\mathrm{d}} \left| a_{kk} \right|^2}{p_{\mathrm{d}} \sum_{i \neq k}^{K} \left| a_{ki} \right|^2 + 1} \right) \right\}. \tag{23}$$

Figure 4 compares the spectral efficiency given by (12), where the kth user estimates the elements of $\boldsymbol{a}_k$ independently, with the one obtained by (23), where we assume that there is a genie receiver at the kth user. Here, we choose $K = 5$ and $T = 200$. We can see that performance gap between two cases is very small. This implies that estimating the elements of $\boldsymbol{a}_k$ independently is fairly reasonable.

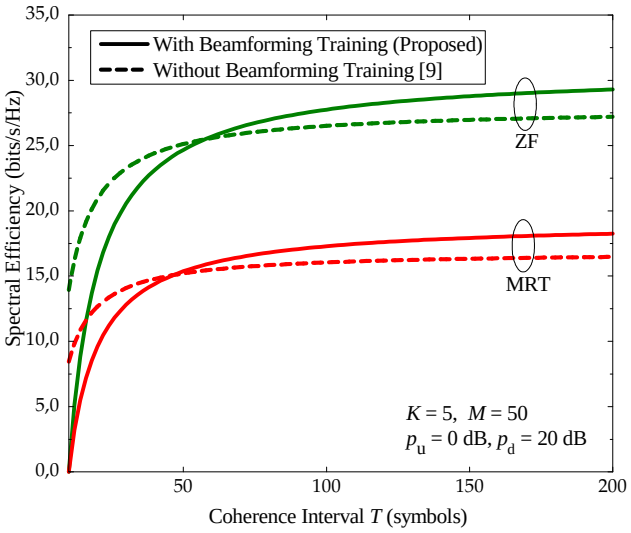


Figure 4: Spectral efficiency versus coherence interval for MRT and ZF precoding ($M = 50$, $K = 5$, $p_{\mathrm{u}} = 0$ dB, and $p_{\mathrm{d}} = 20$ dB).

5 Conclusion and Future Work

In this paper, we proposed and analyzed a scheme to acquire CSI at each user in the downlink of a MU-MIMO system, called beamforming training scheme. With this scheme, the BS uses linear precoding techniques to process the pilot sequence before sending it to the users for the channel estimation. The channel estimation overhead of this beamforming training scheme is small and does not depend on the number of BS antennas. Therefore, it is suitable and efficient for massive MU-MIMO systems. Furthermore, the down-link pilots will add robustness to the beamforming process which otherwise is dependent on the validity of the prior (Bayes) assumptions.

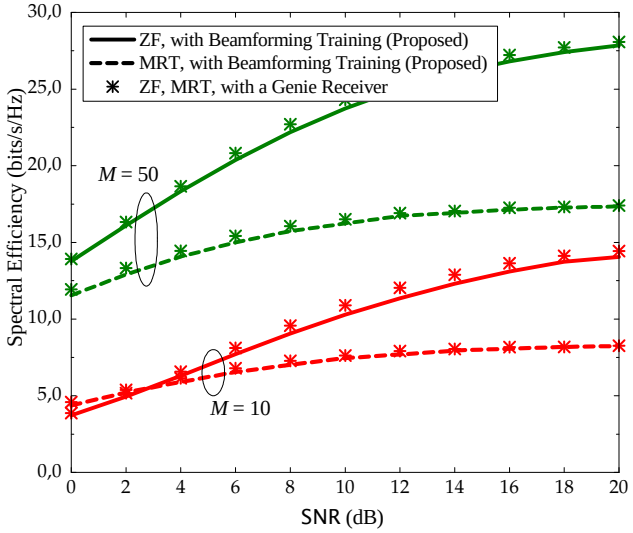


Figure 5: Spectral efficiency versus SNR with a genie receiver ($K = 5$, $p_{\mathrm{u}} = 0$ dB, and $T = 200$).

Appendix

A Proof of Proposition 10

With MRT, we have that $a_{ki} = \alpha_{\texttt{MRT}} \boldsymbol{h}_k^T \hat{\boldsymbol{h}}_i^*$.

- Compute $\mathbb{E}\{a_{ki}\}$:
  From (2), we have

$$a_{ki} = \alpha_{\texttt{MRT}} \left(\hat{\boldsymbol{h}}_k^T + \boldsymbol{\varepsilon}_k^T \right) \hat{\boldsymbol{h}}_i^* = \alpha_{\texttt{MRT}} \hat{\boldsymbol{h}}_k^T \hat{\boldsymbol{h}}_i^* + \alpha_{\texttt{MRT}} \boldsymbol{\varepsilon}_k^T \hat{\boldsymbol{h}}_i^*, \tag{24}$$

  where $\hat{\boldsymbol{h}}_k$ and $\boldsymbol{\varepsilon}_k$ are the kth columns of $\hat{\boldsymbol{H}}$ and $\boldsymbol{\mathcal{E}}$, respectively. Since $\hat{\boldsymbol{\varepsilon}}_k$ and $\hat{\boldsymbol{h}}_i^*$ are uncorrelated with all $i, k = 1, ..., K$, we obtain

$$\mathbb{E}\{a_{ki}\} = \alpha_{\texttt{MRT}} \mathbb{E}\left\{ \hat{\boldsymbol{h}}_k^T \hat{\boldsymbol{h}}_i^* \right\} = \begin{cases} 0, \text{ if } i \neq k \\ \sqrt{\frac{\tau_{\text{u}} p_{\text{u}} M}{K(\tau_{\text{u}} p_{\text{u}} + 1)}}, \text{ if } i = k. \end{cases} \tag{25}$$

- Compute $\texttt{Var}(a_{ki})$ for $i \neq k$:
  From (24) and (25), we have

$$\begin{aligned} \texttt{Var}(a_{ki}) &= \mathbb{E}\left\{ |a_{ki}|^2 \right\} \overset{(a)}{=} \mathbb{E}\left\{ \left| \alpha_{\texttt{MRT}} \hat{\boldsymbol{h}}_k^T \hat{\boldsymbol{h}}_i^* \right|^2 \right\} + \mathbb{E}\left\{ \left| \alpha_{\texttt{MRT}} \boldsymbol{\varepsilon}_k^T \hat{\boldsymbol{h}}_i^* \right|^2 \right\} \\ &= \alpha_{\texttt{MRT}}^2 \left(\frac{\tau_{\text{u}} p_{\text{u}}}{\tau_{\text{u}} p_{\text{u}} + 1} \right)^2 M + \alpha_{\texttt{MRT}}^2 \frac{\tau_{\text{u}} p_{\text{u}} M}{(\tau_{\text{u}} p_{\text{u}} + 1)^2} = 1/K, \end{aligned} \tag{26}$$

  where (a) is obtained by using the fact that $\hat{\boldsymbol{h}}_k^T \hat{\boldsymbol{h}}_i^*$ and $\boldsymbol{\varepsilon}_k^T \hat{\boldsymbol{h}}_i^*$ are uncorrelated.

- Compute $\texttt{Var}(a_{kk})$:
  Similarly, we have

$$\texttt{Var}(a_{kk}) = \mathbb{E}\left\{ |a_{kk}|^2 \right\} - |\mathbb{E}\{a_{kk}\}|^2. \tag{27}$$

From (24), we have

$$\mathbb{E}\left\{|a_{kk}|^2\right\} = \alpha_{\texttt{MRT}}^2 \mathbb{E}\left\{\left\|\hat{\boldsymbol{h}}_k\right\|^4\right\} + \alpha_{\texttt{MRT}}^2 \mathbb{E}\left\{\left|\boldsymbol{\varepsilon}_k^T \hat{\boldsymbol{h}}_k^*\right|^2\right\}. \tag{28}$$

Using [13, Lemma 2.9], we obtain

$$\mathbb{E}\left\{|a_{kk}|^2\right\} = \alpha_{\texttt{MRT}}^2 \left(\frac{\tau_{\mathrm{u}} p_{\mathrm{u}}}{\tau_{\mathrm{u}} p_{\mathrm{u}} + 1}\right)^2 M\left(M+1\right) + \alpha_{\texttt{MRT}}^2 \frac{\tau_{\mathrm{u}} p_{\mathrm{u}}}{\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)^2} M. \tag{29}$$

Substituting (25) and (29) into (27), we obtain

$$\texttt{Var}\left(a_{kk}\right) = 1/K. \tag{30}$$

Substituting (25), (26), and (30) into (10), we get (16).

- Compute $\mathbb{E}\left\{|\epsilon_{ki}|^2\right\}$:

  If $i \neq k$, from (9) and (16), we have

$$\begin{aligned}\mathbb{E}\left\{|\epsilon_{ki}|^2\right\} &= \mathbb{E}\left\{|a_{ki} - \hat{a}_{ki}|^2\right\} \\ &= \mathbb{E}\left\{\left|\frac{K}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K} a_{ki} - \frac{\sqrt{\tau_{\mathrm{d}} p_{\mathrm{d}}}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K} \tilde{n}_{\mathrm{p},ki}\right|^2\right\} \\ &= \left(\frac{K}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K}\right)^2 \mathbb{E}\left\{|a_{ki}|^2\right\} + \frac{\tau_{\mathrm{d}} p_{\mathrm{d}}}{\left(\tau_{\mathrm{d}} p_{\mathrm{d}} + K\right)^2},\end{aligned} \tag{31}$$

  where $\tilde{n}_{\mathrm{p},ki}$ is the ith element of $\tilde{\boldsymbol{n}}_{\mathrm{p},k}$. Using (26), we obtain

$$\mathbb{E}\left\{|\epsilon_{ki}|^2\right\} = \frac{1}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K}. \tag{32}$$

  Similarly, we obtain $\mathbb{E}\left\{|\epsilon_{kk}|^2\right\} = \frac{1}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K}$. Therefore, we arrive at the result in Proposition 10.

B Proof of Proposition 11

With ZF, we have that $a_{ki} = \boldsymbol{h}_k^T \boldsymbol{w}_i$, where $\boldsymbol{w}_i$ is the ith column of $\alpha_{\texttt{ZF}} \hat{\boldsymbol{H}}^* \left(\hat{\boldsymbol{H}}^T \hat{\boldsymbol{H}}^*\right)^{-1}$. Since $\hat{\boldsymbol{H}}^T \boldsymbol{W} = \alpha_{\texttt{ZF}} \boldsymbol{I}_K$, we have

$$a_{ki} = \left(\hat{\boldsymbol{h}}_k^T + \boldsymbol{\varepsilon}_k^T\right) \boldsymbol{w}_i = \alpha_{\texttt{ZF}} \delta_{ki} + \boldsymbol{\varepsilon}_k^T \boldsymbol{w}_i. \tag{33}$$

Therefore,

$$\mathbb{E}\left\{a_{ki}\right\} = \alpha_{\texttt{ZF}} \delta_{ki}. \tag{34}$$

- Compute $\mathtt{Var}\left(a_{ki}\right)$:
  From (33) and (34), we have

$$\begin{aligned}\mathtt{Var}\left(a_{ki}\right) &= \mathbb{E}\left\{\left|\boldsymbol{\epsilon}_k^T \boldsymbol{w}_i\right|^2\right\} = \frac{1}{\tau_{\mathrm{u}} p_{\mathrm{u}} + 1}\,\mathbb{E}\left\{\left\|\boldsymbol{w}_i\right\|^2\right\} \\ &= \frac{\alpha_{\mathtt{ZF}}^2}{\tau_{\mathrm{u}} p_{\mathrm{u}} + 1}\,\mathbb{E}\left\{\left[\left(\hat{\boldsymbol{H}}^T \hat{\boldsymbol{H}}^*\right)^{-1}\right]_{ii}\right\} \\ &= \frac{\alpha_{\mathtt{ZF}}^2}{\tau_{\mathrm{u}} p_{\mathrm{u}} + 1}\frac{1}{K}\,\mathbb{E}\left\{\mathrm{tr}\left[\left(\hat{\boldsymbol{H}}^T \hat{\boldsymbol{H}}^*\right)^{-1}\right]\right\}. \end{aligned} \tag{35}$$

  Using [13, Lemma 2.10], we obtain

$$\mathtt{Var}\left(a_{ki}\right) = \frac{1}{K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}. \tag{36}$$

  Substituting (34) and (36) into (10), we get (20).

- Compute $\mathbb{E}\left\{\left|\epsilon_{ki}\right|^2\right\}$:
  If $i \neq k$, from (9) and (20), we have

$$\begin{aligned}\mathbb{E}\left\{\left|\epsilon_{ki}\right|^2\right\} &= \mathbb{E}\left\{\left|a_{ki} - \hat{a}_{ki}\right|^2\right\} \\ &= \mathbb{E}\left\{\left|\frac{K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right) a_{ki}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)} - \frac{\sqrt{\tau_{\mathrm{d}} p_{\mathrm{d}}}\tilde{n}_{\mathrm{p},ki}}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}\right|^2\right\} \\ &= \left(\frac{K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}\right)^2 \mathbb{E}\left\{\left|a_{ki}\right|^2\right\} + \frac{\tau_{\mathrm{d}} p_{\mathrm{d}}}{\left(\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)\right)^2} \\ &= \frac{1}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}, \end{aligned} \tag{37}$$

  where the last equality is obtained by using (36). Similarly, we obtain $\mathbb{E}\left\{\left|\epsilon_{kk}\right|^2\right\} = \frac{1}{\tau_{\mathrm{d}} p_{\mathrm{d}} + K\left(\tau_{\mathrm{u}} p_{\mathrm{u}} + 1\right)}$. Therefore, we arrive at the result in Proposition 11.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] Y.-H. Nam, B. L. Ng, K. Sayana, Y. Li, J. C. Zhang, Y. Kim, and J. Lee, "Full-dimension MIMO (FD-MIMO) for next generation cellular technology," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 172–178, 2013.

[4] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, Istanbul, Turkey, Aug. 2012.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[6] 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.

[7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[8] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[9] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[10] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Hardware impairments in large-scale MISO systems: Energy efficiency, estimation, and capacity limits," in *Proc. Signal Processing and Optimization for Green Energy and Green Communications (DSP'13)*, Santorini, Greece, 2013.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[12] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, June 2013.

[13] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

Paper F

Blind Estimation of Effective Downlink Channel Gains in Massive MIMO

Refereed article submitted to the IEEE ICASSP 2015.

Blind Estimation of Effective Downlink Channel Gains in Massive MIMO

Hien Quoc Ngo and Erik G. Larsson

Abstract

We consider the massive MIMO downlink with time-division duplex (TDD) operation and conjugate beamforming transmission. To reliably decode the desired signals, the users need to know the effective channel gain. In this paper, we propose a blind channel estimation method which can be applied at the users and which does not require any downlink pilots. We show that our proposed scheme can substantially outperform the case where each user has only statistical channel knowledge, and that the difference in performance is particularly large in certain types of channel, most notably keyhole channels. Compared to schemes that rely on downlink pilots (e.g., [1]), our proposed scheme yields more accurate channel estimates for a wide range of signal-to-noise ratios and avoid spending time-frequency resources on pilots.

1 Introduction

Massive multiple-input multiple-output (MIMO) is one of the most promising technologies to meet the demands for high throughput and communication reliability of next generation cellular networks [2–5]. In massive MIMO, time-division duplex (TDD) operation is preferable since then the pilot overhead does not depend on the number of base station antennas. With TDD, the channels are estimated at the base station through the uplink training. For the downlink, under the assumption of channel reciprocity, the channels estimated at the base station are used to precode the data, and the precoded data are sent to the users. To coherently decode the transmitted signals, each user should have channel state information (CSI), that is, know its effective channel from the base station.

In most previous works, the users are assumed to have statistical knowledge of the effective downlink channels, that is, they know the mean of the effective channel gain and use this for the signal detection [6, 7]. In these papers, Rayleigh fading channels were assumed. Under the Rayleigh fading, the effective channel gains become nearly deterministic (the channel "hardens") when the number of base station antennas grows large, and hence, using the mean of the effective channel gain for signal detection works very well. However, in practice, propagation scenarios may be encountered where the channel does not harden. In that case, using the mean effective channel gain may not be accurate enough, and a better estimate of the effective channel should be used. In [1], we proposed a scheme where the base station (in addition to the beamformed data) also sent a beamformed downlink pilot sequence to the users. With this scheme, a performance improvement (compared to the case when the *mean* of the effective channel gain is used) was obtained. However, this scheme requires time-frequency resources in order to send the downlink pilots. The associated overhead is proportional to the number of users which can be in the order of several tens, and hence, in a high-mobility environment (where the channel coherence interval is short) the spectral efficiency is significantly reduced.

Contribution: In this paper, we consider the massive MIMO downlink with conjugate beamforming.[1] We propose a scheme with which the users *blindly* estimate the effective channel gain from the received data. The scheme exploits the asymptotic properties of the mean of the received signal power when the number of base station antennas is large. The accuracy of our proposed scheme is investigated for two specific, very different, types of channels: (i) independent Rayleigh fading and (ii) keyhole channels. We show that when the number of base station antennas goes to infinity, the channel estimate provided by our scheme becomes exact. Also, numerical results quantitatively show the benefits of our proposed scheme, especially in keyhole channels, compared to the case where the mean of the effective channel

[1] We consider conjugate beamforming since it is simple and nearly optimal in many massive MIMO scenarios. More importantly, conjugate beamforming can be implemented in a distributed manner.

gain is used as if it were the true channel gain, and compared to the case where the beamforming training scheme of [1] is used.

Notation: We use boldface upper- and lower-case letters to denote matrices and column vectors, respectively. The superscripts $()^T$ and $()^H$ stand for the transpose and conjugate transpose, respectively. The Euclidean norm, the trace, and the expectation operators are denoted by $\|\cdot\|$, $\mathrm{Tr}\,(\cdot)$, and $\mathbb{E}\{\cdot\}$, respectively. The notation $\xrightarrow{P}$ means convergence in probability, and $\xrightarrow{a.s.}$ means almost sure convergence. Finally, we use $z \sim \mathcal{CN}\left(0, \sigma^2\right)$ to denote a circularly symmetric complex Gaussian random variable (RV) z with zero mean and variance σ^2.

2 System Model

Consider the downlink of a massive MIMO system. An M-antenna base station serves K single-antenna users, where $M \gg K \gg 1$. The base station uses conjugate beamforming to simultaneously transmit data to all K users in the same time-frequency resource. Since we focus on the downlink channel estimation here, we assume that the base station perfectly estimates the channels in the uplink training phase. (In future work, this assumption may be relaxed.) Denote by $\boldsymbol{g}_k$ the $M \times 1$ channel vector between the base station and the kth user. The channel $\boldsymbol{g}_k$ results from a combination of small-scale fading and large-scale fading, and is modeled as:

$$\boldsymbol{g}_k = \sqrt{\beta_k}\boldsymbol{h}_k, \tag{1}$$

where β_k represents large-scale fading which is constant over many coherence intervals, and $\boldsymbol{h}_k$ is an $M \times 1$ small-scale channel vector. We assume that the elements of $\boldsymbol{h}_k$ are i.i.d. with zero mean and unit variance.

Let s_k, $\mathbb{E}\left\{|s_k|^2\right\} = 1$, $k = 1, \ldots, K$, be the symbol intended for the kth user. With conjugate beamforming, the $M \times 1$ precoded signal vector is given by

$$\boldsymbol{x} = \sqrt{\alpha}\boldsymbol{G}\boldsymbol{s}, \tag{2}$$

where $\boldsymbol{s} \triangleq [s_1, s_2, \ldots, s_K]^T$, $\boldsymbol{G} \triangleq [\boldsymbol{g}_1 \ldots \boldsymbol{g}_K]$ is an $M \times K$ channel matrix between the K users and the base station, and α is a normalization constant chosen to satisfy the average power constraint at the base station:

$$\mathbb{E}\left\{\|\boldsymbol{x}\|^2\right\} = \rho.$$

Hence,

$$\alpha = \frac{\rho}{\mathbb{E}\left\{\mathrm{Tr}\left(\boldsymbol{G}\boldsymbol{G}^H\right)\right\}}. \tag{3}$$

The signal received at the kth user is

$$\begin{aligned} y_k &= \boldsymbol{g}_k^H \boldsymbol{x} + n_k = \sqrt{\alpha}\boldsymbol{g}_k^H \boldsymbol{G}\boldsymbol{s} + n_k \\ &= \sqrt{\alpha}\|\boldsymbol{g}_k\|^2 s_k + \sqrt{\alpha}\sum_{k' \neq k}^{K} \boldsymbol{g}_k^H \boldsymbol{g}_{k'} s_{k'} + n_k, \end{aligned} \tag{4}$$

where $n_k \sim \mathcal{CN}(0,1)$ is the additive Gaussian noise at the kth user. Then, the desired signal s_k is decoded.

3 Proposed Downlink Blind Channel Estimation Technique

The kth user wants to detect s_k from y_k in (4). For this purpose, it needs to know the effective channel gain $\|\boldsymbol{g}_k\|^2$. If the channel is Rayleigh fading, then by the law of large numbers, we have

$$\frac{1}{M}\|\boldsymbol{g}_k\|^2 \xrightarrow{P} \beta_k,$$

as $M \to \infty$. This implies that when M is large, $\|\boldsymbol{g}_k\|^2 \approx M\beta_k$ (we say that the channel *hardens*). So we can use the statistical properties of the channel, i.e., use $\mathbb{E}\left\{\|\boldsymbol{g}_k\|^2\right\} = M\beta_k$ as a good estimate of $\|\boldsymbol{g}_k\|^2$ when detecting s_k. This assumption is widely made in the massive MIMO literature. However, in practice, the channel is not always Rayleigh fading, and does not always harden when $M \to \infty$. For example, consider a keyhole channel, where the small-scale fading component $\boldsymbol{h}_k$ is modeled as follows [8, 9]:

$$\boldsymbol{h}_k = \nu_k \bar{\boldsymbol{h}}_k, \tag{5}$$

where ν_k and the M elements of $\bar{\boldsymbol{h}}_k$ are i.i.d. $\mathcal{CN}(0,1)$ RVs. For the keyhole channel (5), by the law of large numbers, we have

$$\frac{1}{M}\|\boldsymbol{g}_k\|^2 - \beta_k|\nu_k|^2 \xrightarrow{P} 0,$$

which is not deterministic, and hence the channel does not harden. In this case, using $\mathbb{E}\left\{\|\boldsymbol{g}_k\|^2\right\} = M\beta_k$ as an estimate of the true effective channel $\|\boldsymbol{g}_k\|^2$ to detect s_k may result in poor performance.

For the reasons explained, it is desirable that the users estimate their effective channels. One way to do this is to have the base station transmit beamformed downlink pilots as proposed in [1]. With this scheme, at least K downlink pilot symbols are required. This can significantly reduce the spectral efficiency. For example, suppose $M = 300$ antennas serve $K = 50$ terminals, in a coherence interval of length 200 symbols. If half of the coherence interval is used for the downlink,

then with the downlink beamforming training of [1], we need to spend at least 50 symbols for sending pilots. As a result, less than 50 of the 100 downlink symbols are used for payload in each coherence interval, and the insertion of the downlink pilots reduces the overall (uplink+downlink) spectral efficiency by a factor of 1/4.

In what follows, we propose a blind channel estimation method which does not require any downlink pilots.

3.1 Mathematical Preliminaries

Consider the average power of the received signal at the kth user (averaged over $\boldsymbol{s}$ and n_k). From (4), we have

$$\mathbb{E}\left\{|y_k|^2\right\} = \alpha\|\boldsymbol{g}_k\|^4 + \alpha\sum_{k'\neq k}^{K}\left|\boldsymbol{g}_k^H\boldsymbol{g}_{k'}\right|^2 + 1. \tag{6}$$

The second term of (6) can be rewritten as

$$\alpha\sum_{k'\neq k}^{K}\left|\boldsymbol{g}_k^H\boldsymbol{g}_{k'}\right|^2 = \alpha\sum_{k'\neq k}^{K}\boldsymbol{g}_{k'}^H\boldsymbol{g}_k\boldsymbol{g}_k^H\boldsymbol{g}_{k'} = \alpha\tilde{\boldsymbol{g}}_k^H\boldsymbol{A}\tilde{\boldsymbol{g}}_k, \tag{7}$$

where $\tilde{\boldsymbol{g}}_k \triangleq [\boldsymbol{g}_1^T \ \cdots \ \boldsymbol{g}_{k-1}^T \ \boldsymbol{g}_{k+1}^T \ \cdots \ \boldsymbol{g}_K^T]^T$, and $\boldsymbol{A}$ is an $M(K-1)\times M(K-1)$ block-diagonal matrix whose (i,i)-block is the $M\times M$ matrix $\boldsymbol{g}_k\boldsymbol{g}_k^H$. Since $\boldsymbol{A}$ and $\tilde{\boldsymbol{g}}_k$ are independent, as $M(K-1)\to\infty$, the Trace lemma gives [10]

$$\frac{1}{M(K-1)}\sum_{k'\neq k}^{K}\left|\boldsymbol{g}_k^H\boldsymbol{g}_{k'}\right|^2 - \frac{1}{M(K-1)}\sum_{k'\neq k}^{K}\beta_{k'}\|\boldsymbol{g}_k\|^2 \overset{a.s.}{\to} 0. \tag{8}$$

Substituting (8) into (6), as $M(K-1)\to\infty$, we have

$$\frac{\mathbb{E}\left\{|y_k|^2\right\}}{M(K-1)} - \frac{1}{M(K-1)}\left(\alpha\|\boldsymbol{g}_k\|^4 + \alpha\sum_{k'\neq k}^{K}\beta_{k'}\|\boldsymbol{g}_k\|^2 + 1\right) \overset{a.s.}{\to} 0. \tag{9}$$

The above result implies that when M and K are large,

$$\mathbb{E}\left\{|y_k|^2\right\} \approx \alpha\|\boldsymbol{g}_k\|^4 + \alpha\sum_{k'\neq k}^{K}\beta_{k'}\|\boldsymbol{g}_k\|^2 + 1. \tag{10}$$

Therefore, the effective channel gain $\|\boldsymbol{g}_k\|^2$ can be estimated from $\mathbb{E}\left\{|y_k|^2\right\}$ by solving the quadratic equation (10).

3.2 Downlink Blind Channel Estimation Algorithm

As discussed in Section 3.1, we can estimate the effective channel gain $\|\boldsymbol{g}_k\|^2$ by solving the quadratic equation (10). It is then required that the kth user knows α, $\sum_{k'\neq k}^{K}\beta_{k'}$, and $\mathbb{E}\left\{|y_k|^2\right\}$. We assume that the kth user knows α and $\sum_{k'\neq k}^{K}\beta_{k'}$. This assumption is reasonable since the terms α and $\sum_{k'\neq k}^{K}\beta_{k'}$ depend on the large-scale fading coefficients, which stay constant over many coherence intervals. The kth user can estimate these terms, or the base station may inform the kth user about them. Regarding $\mathbb{E}\left\{|y_k|^2\right\}$, in practice, it is unavailable. However, we can use the received samples during a whole coherence interval to form a sample estimate of $\mathbb{E}\left\{|y_k|^2\right\}$ as follows:

$$\mathbb{E}\left\{|y_k|^2\right\} \approx \xi_k \triangleq \frac{|y_k(1)|^2+|y_k(2)|^2+\ldots+|y_k(T)|^2}{T}, \tag{11}$$

where $y_k(n)$ is the nth receive sample, and T is the length (in symbols) of the coherence interval used for the downlink transmission.

The algorithm for estimating $\|\boldsymbol{g}_k\|^2$ is summarized as follows:

Algorithm 3 *(Proposed blind downlink channel estimation method)*

1. *Using a data block of T samples, compute ξ_k as* (11).
2. *The channel estimate of $\|\boldsymbol{g}_k\|^2$, denoted by a_k, is determined as*

$$a_k = \frac{-\alpha\sum_{k'\neq k}^{K}\beta_{k'}+\sqrt{\alpha^2\left(\sum_{k'\neq k}^{K}\beta_{k'}\right)^2+4\alpha(\xi_k-1)}}{2\alpha}. \tag{12}$$

Note that a_k in (12) is the positive root of the quadratic equation: $\xi_k = \alpha a_k^2 + \alpha\sum_{k'\neq k}^{K}\beta_{k'}a_k + 1$ which comes from (10) and (11).

3.3 Asymptotic Performance Analysis

In this section, we analyze the accuracy of our proposed scheme for two specific propagation environments: Rayleigh fading and keyhole channels. For keyhole channels, we use the model (5). We assume that the kth user perfectly estimates $\mathbb{E}\left\{|y_k|^2\right\}$. This is true when the number of symbols of the coherence interval allocated to the downlink, T, is large. In the numerical results, we shall show that the estimate of $\mathbb{E}\left\{|y_k|^2\right\}$ in (11) is very close to $\mathbb{E}\left\{|y_k|^2\right\}$ even for modest values of T (e.g. $T \approx 100$

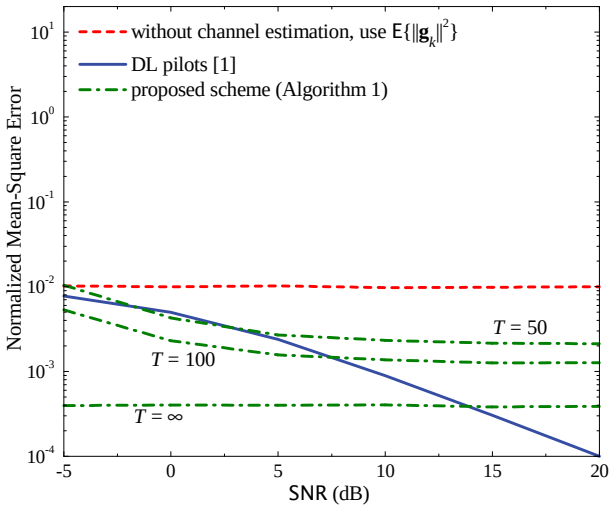


Figure 1: Normalized MSE versus SNR for different channel estimation schemes, for Rayleigh fading channels.

symbols). With the assumption $\xi_k = \mathbb{E}\left\{|y_k|^2\right\}$, from (6) and (12), the estimate of $\|\boldsymbol{g}_k\|^2$ can be written as:

$$a_k = -\frac{\sum_{k'\neq k}^{K}\beta_{k'}}{2} + \sqrt{\left(\frac{\sum_{k'\neq k}^{K}\beta_{k'}}{2} + \|\boldsymbol{g}_k\|^2\right)^2 + \epsilon_k}, \tag{13}$$

where

$$\epsilon_k \triangleq \sum_{k'\neq k}^{K}\left|\boldsymbol{g}_k^H\boldsymbol{g}_{k'}\right|^2 - \left(\sum_{k'\neq k}^{K}\beta_{k'}\right)\|\boldsymbol{g}_k\|^2. \tag{14}$$

We can see from (13) that if $|\epsilon_k| \ll \left(\frac{\sum_{k'\neq k}^{K}\beta_{k'}}{2} + \|\boldsymbol{g}_k\|^2\right)^2$, then $a_k \approx \|\boldsymbol{g}_k\|^2$. In order to see under what conditions $|\epsilon_k| \ll \left(\frac{\sum_{k'\neq k}^{K}\beta_{k'}}{2} + \|\boldsymbol{g}_k\|^2\right)^2$, we consider ϱ_k which is defined as:

$$\varrho_k \triangleq \mathbb{E}\left\{\left|\epsilon_k / \mathbb{E}\left\{\left(\frac{1}{2}\sum_{k'\neq k}^{K}\beta_{k'} + \|\boldsymbol{g}_k\|^2\right)^2\right\}\right|^2\right\}. \tag{15}$$

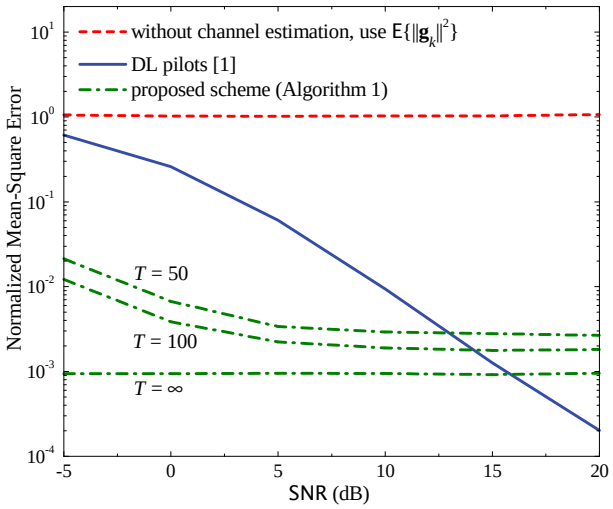


Figure 2: Normalized MSE versus SNR for different channel estimation schemes, for keyhole channels.

Hence,

$$\varrho_k = \begin{cases} \dfrac{M(M+1)\beta_k^2 \sum\limits_{k'\neq k}^{K} \beta_{k'}^2}{\left(\frac{1}{4}\bar{\beta}_k^2 + M\beta_k \sum\limits_{k'=1}^{K} \beta_{k'} + \beta_k^2 M^2\right)^2}, & \text{for Rayleigh fading channels,} \\ \dfrac{6M(M+1)\beta_k^2 \sum\limits_{k'\neq k}^{K} \beta_{k'}^2}{\left(\frac{1}{4}\bar{\beta}_k^2 + M\beta_k \sum\limits_{k'=1}^{K} \beta_{k'} + \beta_k^2 M(2M+1)\right)^2}, & \text{for keyhole channels,} \end{cases} \tag{16}$$

where $\bar{\beta}_k \triangleq \sum_{k'\neq k}^{K} \beta_{k'}$. The detailed derivations of (16) are presented in the Appendix. We can see that $\varrho_k = O(1/M^2)$. Thus, when $M \gg 1$, $|\epsilon_k|$ is much smaller than $\left(\frac{\sum_{k'\neq k}^{K} \beta_{k'}}{2} + \|\boldsymbol{g}_k\|^2\right)^2$. As a result, our proposed channel estimation scheme is expected to work well.

4 Numerical Results

In this section, we provide numerical results to evaluate our proposed channel estimation scheme for finite M. As performance metric we consider the normalized mean-square error (MSE) at the kth user:

$$\mathtt{MSE}_k \triangleq \mathbb{E}\left\{ \left| \frac{a_k - \|\boldsymbol{g}_k\|^2}{\mathbb{E}\{\|\boldsymbol{g}_k\|^2\}} \right|^2 \right\}. \tag{17}$$

For the simulation, we choose $M = 100$, $K = 20$, and $\beta_k = 1, \forall k = 1, \ldots, K$. We define $\mathsf{SNR} \triangleq \rho$. Figures 1 and 2 show the normalized MSE versus SNR for Rayleigh fading and keyhole channels, respectively. The curves labeled "without channel estimation, use $\mathbb{E}\left\{\|\boldsymbol{g}_k\|^2\right\}$" represent the case when the kth user uses the statistical properties of the channels, i.e., it uses $\mathbb{E}\left\{\|\boldsymbol{g}_k\|^2\right\}$ as estimate of $\|\boldsymbol{g}_k\|^2$. The curves "DL pilots [1]" represent the case when the beamforming training scheme of [1] with MMSE channel estimation is applied. The curves "proposed scheme (Algorithm 1)" represent our proposed scheme for different T ($T = \infty$ implies that the kth user perfectly knows $\mathbb{E}\left\{|y_k|^2\right\}$). For the beamforming training scheme, the duration of the downlink training is K. For our proposed blind channel estimation scheme, $s_k, k = 1, \ldots, K$, are random 4-QAM symbols.

We can see that in Rayleigh fading channels, the MSEs of the three schemes are comparable. Using $\mathbb{E}\left\{\|\boldsymbol{g}_k\|^2\right\}$ in lieu of the true $\|\boldsymbol{g}_k\|^2$ for signal detection works rather well. However, in keyhole channels, since the channels do not harden, the MSE when using $\mathbb{E}\left\{\|\boldsymbol{g}_k\|^2\right\}$ as estimate of $\|\boldsymbol{g}_k\|^2$ is very large. In both propagation environments, our proposed scheme works very well. For a wide range of SNRs, our scheme outperforms the beamforming training scheme, even for short coherence intervals (e.g., $T = 100$ symbols). Note again that, with the beamforming training scheme of [1], we additionally have to spend at least K symbols on training pilots (this is not accounted for here, since we only evaluated MSE). By contrast, our proposed scheme does not requires any resources for downlink training.

5 Concluding Remarks

Massive MIMO systems may encounter propagation conditions when the channels do not harden. Then, to facilitate detection of the data in the downlink, the users need to estimate their effective channel gain rather than relying on knowledge of the *average* effective channel gain. We proposed a channel estimation approach by which the users can blindly estimate the effective channel gain from the data received during a coherence interval. The approach is computationally easy, it does not requires any resource for downlink pilots, it can be applied regardless of the type of propagation channel, and it performs very well.

Future work may include studying rate expressions rather than channel estimation MSE, and taking into account the channel estimation errors in the uplink. (We hypothesize, that the latter will not qualitatively affect our results or conclusions.) Blind estimation of β_k by the users may also be addressed.

Appendix

Here, we provide the proof of (16). From (15), we have

$$\varrho_k = \mathbb{E}\left\{|\epsilon_k|^2\right\} / \mathbb{E}\left\{\left(\frac{1}{2}\sum_{k'\neq k}^{K}\beta_{k'} + \|\boldsymbol{g}_k\|^2\right)^2\right\}. \tag{18}$$

- *Rayleigh Fading Channels:*

For Rayleigh fading channels, we have

$$\begin{aligned}\mathbb{E}\left\{\left(\frac{1}{2}\sum_{k'\neq k}^{K}\beta_{k'} + \|\boldsymbol{g}_k\|^2\right)^2\right\} &= \frac{1}{4}\left(\sum_{k'\neq k}^{K}\beta_{k'}\right)^2 + \left(\sum_{k'\neq k}^{K}\beta_{k'}\right)\mathbb{E}\left\{\|\boldsymbol{g}_k\|^2\right\} + \mathbb{E}\left\{\|\boldsymbol{g}_k\|^4\right\} \\ &= \frac{1}{4}\left(\sum_{k'\neq k}^{K}\beta_{k'}\right)^2 + M\beta_k\sum_{k'=1}^{K}\beta_{k'} + \beta_k^2 M^2, \end{aligned} \tag{19}$$

where the last equality follows [11, Lemma 2.9]. We next compute $\mathbb{E}\left\{|\epsilon_k|^2\right\}$. From (14), we have

$$\begin{aligned}\mathbb{E}\left\{|\epsilon_k|^2\right\} = \mathbb{E}&\left\{\left(\sum_{k'\neq k}^{K}\left|\boldsymbol{g}_k^H\boldsymbol{g}_{k'}\right|^2\right)^2\right\} + \left(\sum_{k'\neq k}^{K}\beta_{k'}\right)^2\mathbb{E}\left\{\|\boldsymbol{g}_k\|^4\right\} \\ &- 2\left(\sum_{k'\neq k}^{K}\beta_{k'}\right)\mathbb{E}\left\{\sum_{k'\neq k}^{K}\left|\boldsymbol{g}_k^H\boldsymbol{g}_{k'}\right|^2\|\boldsymbol{g}_k\|^2\right\}. \end{aligned} \tag{20}$$

We have,

$$\mathbb{E}\left\{\left(\sum_{k'\neq k}^{K}\left|\boldsymbol{g}_k^H\boldsymbol{g}_{k'}\right|^2\right)^2\right\} = \mathbb{E}\left\{\|\boldsymbol{g}_k\|^4\left(\sum_{k'\neq k}^{K}|z_{k'}|^2\right)^2\right\}, \tag{21}$$

where $z_{k'} \triangleq \frac{\boldsymbol{g}_k^H \boldsymbol{g}_{k'}}{\|\boldsymbol{g}_k\|}$. Conditioned on $\boldsymbol{g}_k$, $z_{k'}$ is complex Gaussian distributed with zero mean and variance $\beta_{k'}$ which is independent of $\boldsymbol{g}_k$. Thus, $z_{k'} \sim \mathcal{CN}(0, \beta_{k'})$ and is independent of $\boldsymbol{g}_k$. This yields

$$\mathbb{E}\left\{\left(\sum_{k'\neq k}^{K} \left|\boldsymbol{g}_k^H \boldsymbol{g}_{k'}\right|^2\right)^2\right\} = \mathbb{E}\left\{\|\boldsymbol{g}_k\|^4\right\} \mathbb{E}\left\{\left(\sum_{k'\neq k}^{K} |z_{k'}|^2\right)^2\right\}$$
$$= \beta_k^2 M(M+1)\left(\sum_{i\neq k}^{K} \beta_i^2 + \sum_{i\neq k}^{K}\sum_{j\neq k}^{K} \beta_i \beta_j\right). \quad (22)$$

Similarly,

$$\mathbb{E}\left\{\sum_{k'\neq k}^{K} \left|\boldsymbol{g}_k^H \boldsymbol{g}_{k'}\right|^2 \|\boldsymbol{g}_k\|^2\right\} = \mathbb{E}\left\{\|\boldsymbol{g}_k\|^4\right\} \mathbb{E}\left\{\sum_{k'\neq k}^{K} |z_{k'}|^2\right\}$$
$$= \beta_k^2 M(M+1) \sum_{k'\neq k}^{K} \beta_{k'}^2. \quad (23)$$

Substituting (22), (23), and $\mathbb{E}\left\{\|\boldsymbol{g}_k\|^4\right\} = \beta_k^2 M(M+1)$ into (20), we obtain

$$\mathbb{E}\left\{|\epsilon_k|^2\right\} = M(M+1)\beta_k^2 \sum_{k'\neq k}^{K} \beta_{k'}^2. \quad (24)$$

Inserting (19) and (24) into (18), we obtain (16) for the Rayleigh fading case.

- *Keyhole Channels:*

By using the fact that

$$z_{k'} = \frac{\boldsymbol{g}_k^H \boldsymbol{g}_{k'}}{\|\boldsymbol{g}_k\|} = \sqrt{\beta_{k'}}\nu_{k'} \frac{\boldsymbol{g}_k^H \bar{\boldsymbol{h}}_{k'}}{\|\boldsymbol{g}_k\|} \quad (25)$$

is the product of two independent Gaussian RVs, and following a similar methodology used in the Rayleigh fading case, we obtain (16) for keyhole channels.

References

[1] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Proc. Allerton Conference on Communication, Control, and Computing*, Illinois, Oct. 2013.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195 , Feb. 2014.

[3] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.

[4] A. Liu and V. K.N. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, Sept. 2014.

[5] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[6] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[7] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[8] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.

[9] C. Zhong, S. Jin, K.-K. Wong, and M. R. McKay, "Ergodic mutual information analysis for multi-keyhole MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, p. 1754–1763, Jun. 2011.

[10] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Info. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012

[11] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

Paper G

Massive MIMO with Optimal Power and Training Duration Allocation

Refereed article published in the IEEE Wireless
Communications Letters 2014.

Massive MIMO with Optimal Power and Training Duration Allocation

Hien Quoc Ngo, Michail Matthaiou, and Erik G. Larsson

Abstract

We consider the uplink of massive multicell multiple-input multiple-output systems, where the base stations (BSs), equipped with massive arrays, serve simultaneously several terminals in the same frequency band. We assume that the BS estimates the channel from uplink training, and then uses the maximum ratio combining technique to detect the signals transmitted from all terminals in its own cell. We propose an optimal resource allocation scheme which jointly selects the training duration, training signal power, and data signal power in order to maximize the sum spectral efficiency, for a given total energy budget spent in a coherence interval. Numerical results verify the benefits of the optimal resource allocation scheme. Furthermore, we show that more training signal power should be used at low signal-to-noise ratio (SNRs), and vice versa at high SNRs. Interestingly, for the entire SNR regime, the optimal training duration is equal to the number of terminals.

1 Introduction

Massive multiple-input multiple-output (MIMO) has attracted a lot of research interest recently [1–4]. Typically, the uplink transmission in massive MIMO systems consists of two phases: uplink training (to estimate the channels) and uplink payload data transmission. In previous works on massive MIMO, the transmit power of each symbol is assumed to be the same during the training and data transmission phases [1, 5]. However, this equal power allocation policy causes a "squaring effect" in the low power regime [6]. The squaring effect comes from the fact that when the transmit power is reduced, both the data signal and the pilot signal suffer from a power reduction. As a result, in the low power regime, the capacity scales as p_{u}^2, where p_{u} is the transmit power.

In this paper, we consider the uplink of massive multicell MIMO with maximum ratio combining (MRC) receivers at the base station (BS). We consider MRC receivers since they are simple and perform rather well in massive MIMO, particularly when the inherent effect of channel estimation on intercell interference is taken into account [5]. Contrary to most prior works, we assume that the average transmit powers of pilot symbol and data symbol are different. We investigate a resource allocation problem which finds the transmit pilot power, transmit data power, as well as, the training duration that maximize the sum spectral efficiency for a given total energy budget spent in a coherence interval. Our numerical results show appreciable benefits of the proposed optimal resource allocation. At low signal-to-noise ratios (SNRs), more power is needed for training to reduce the squaring effect, while at high SNRs, more power is allocated to data transmission.

Regarding related works, [6–8] elaborated on a similar issue. In [6, 7], the authors considered point-to-point MIMO systems, and in [8], the authors considered single-input multiple-output multiple access channels with scheduling. Most importantly, the performance metric used in [6–8] was the mutual information without any specific signal processing. In this work, however, we consider massive multicell multiuser MIMO systems with MRC receivers and demonstrate the strong potential of these configurations.

2 Massive Multicell MIMO System Model

We consider the uplink multicell MIMO system described in [5]. The system has L cells. Each cell includes one N-antenna BS, and K single-antenna terminals, where $N \gg K$. All L cells share the same frequency band. The transmission comprises two phases: uplink training and data transmission.

2.1 Uplink Training

In the uplink training phase, the BS estimates the channel from received pilot signals transmitted from all terminals. In each cell, K terminals are assigned K orthogonal pilot sequences of length τ symbols ($K \leq \tau \leq T$), where T is the length of the coherence interval. Since the coherence interval is limited, we assume that the same orthogonal pilot sequences are reused in all L cells. This causes the so-called *pilot contamination* [1]. Note that interference from data symbols is as bad as interference from pilots [5].

We denote by $\boldsymbol{G}_{\ell i} \in \mathbb{C}^{N \times K}$ the channel matrix between the BS in the ℓth cell and the K terminals in the ith cell. The (m,k)th element of $\boldsymbol{G}_{\ell i}$ is modeled as

$$g_{\ell imk} = h_{\ell imk}\sqrt{\beta_{\ell ik}}, \quad m = 1, 2, ..., N, \tag{1}$$

where $h_{\ell imk} \sim \mathcal{CN}(0,1)$ represents the small-scale fading coefficient from the mth antenna of the ℓth BS to the kth terminal in the ith cell, and $\sqrt{\beta_{\ell ik}}$ is a constant that represents large-scale fading (pathloss and shadow fading).

At the ℓth BS, the minimum mean-square error channel estimate for the kth column of the channel matrix $\boldsymbol{G}_{\ell\ell}$ is [5]

$$\hat{\boldsymbol{g}}_{\ell\ell k} = \beta_{\ell\ell k}\left(\sum_{j=1}^{L}\beta_{\ell jk} + \frac{1}{\tau p_{\mathrm{p}}}\right)^{-1}\left(\sum_{j=1}^{L}\boldsymbol{g}_{\ell jk} + \frac{\boldsymbol{w}_{\ell k}}{\sqrt{\tau p_{\mathrm{p}}}}\right), \tag{2}$$

where p_{p} is the transmit power of each pilot symbol, and $\boldsymbol{w}_{\ell k} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$ represents additive noise.

2.2 Data Transmission

In this phase, all K terminals send their data to the BS. Let $\sqrt{p_{\mathrm{u}}}\boldsymbol{x}_i \in \mathbb{C}^{K\times 1}$ be a vector of symbols transmitted from the K terminals in the ith cell, where $\mathbb{E}\{\boldsymbol{x}_i\boldsymbol{x}_i^H\} = \boldsymbol{I}_K$, $\mathbb{E}\{\cdot\}$ denotes expectation, and p_{u} be the average transmitted power of each terminal. The $N \times 1$ received vector at the ℓth BS is given by

$$\boldsymbol{y}_\ell = \sqrt{p_{\mathrm{u}}}\sum_{i=1}^{L}\boldsymbol{G}_{\ell i}\boldsymbol{x}_i + \boldsymbol{n}_\ell, \tag{3}$$

where $\boldsymbol{n}_\ell \in \mathbb{C}^{N\times 1}$ is the AWGN vector, distributed as $\boldsymbol{n}_\ell \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$. Then, BS ℓ uses MRC together with the channel estimate to detect the signals transmitted from the K terminals in its own cell. More precisely, to detect the signal transmitted

from the kth terminal, $x_{\ell k}$, the received vector $\boldsymbol{y}_\ell$ is pre-multiplied with $\hat{\boldsymbol{g}}_{\ell\ell k}^H$ to obtain:

$$r_k \triangleq \hat{\boldsymbol{g}}_{\ell\ell k}^H \boldsymbol{y}_\ell = \sqrt{p_{\mathrm{u}}} \hat{\boldsymbol{g}}_{\ell\ell k}^H \boldsymbol{g}_{llk} x_{\ell k} + \sqrt{p_{\mathrm{u}}} \sum_{j \neq k}^{K} \hat{\boldsymbol{g}}_{\ell\ell k}^H \boldsymbol{g}_{\ell\ell j} x_{\ell j} + \sqrt{p_{\mathrm{u}}} \sum_{i \neq \ell}^{L} \hat{\boldsymbol{g}}_{\ell\ell k}^H \boldsymbol{G}_{\ell i} \boldsymbol{x}_i + \hat{\boldsymbol{g}}_{\ell\ell k}^H \boldsymbol{n}_\ell, \tag{4}$$

and then $x_{\ell k}$ can be extracted directly from r_k.

2.3 Sum Spectral Efficiency

In our analysis, the performance metric is the sum spectral efficiency (in bits/s/Hz). From (4), and following a similar methodology as in [5], we obtain an achievable ergodic rate of the transmission from the kth terminal in the ℓth cell to its BS as:[1]

$$R_{\ell k} = \log_2 \left(1 + \frac{a_k \tau p_{\mathrm{p}} p_{\mathrm{u}}}{b_k \tau p_{\mathrm{p}} p_{\mathrm{u}} + c_k p_{\mathrm{u}} + d_k \tau p_{\mathrm{p}} + 1} \right), \tag{5}$$

where $a_k \triangleq \beta_{\ell\ell k}^2 (N-1)$,

$$b_k \triangleq (N-1) \sum_{i \neq \ell}^{L} \beta_{\ell i k}^2 - \sum_{i=1}^{L} \beta_{\ell i k}^2 + \left(\sum_{i=1}^{L} \sum_{j=1}^{K} \beta_{\ell i j} \right) \sum_{i=1}^{L} \beta_{\ell i k},$$

$$c_k \triangleq \sum_{i=1}^{L} \sum_{j=1}^{K} \beta_{\ell i j}, \text{ and } d_k \triangleq \sum_{i=1}^{L} \beta_{\ell i k}.$$

The sum spectral efficiency is defined as

$$\mathcal{S} \triangleq \left(1 - \frac{\tau}{T} \right) \sum_{k=1}^{K} R_{\ell k}. \tag{6}$$

For $p_{\mathrm{u}} \ll 1$, and for p_{p} fixed regardless of p_{u}, we have

$$\mathcal{S} = \log_2 e \left(1 - \frac{\tau}{T} \right) \sum_{k=1}^{K} \frac{a_k \tau p_{\mathrm{p}}}{d_k \tau p_{\mathrm{p}} + 1} p_{\mathrm{u}} + \mathcal{O}\left(p_{\mathrm{u}}^2 \right), \tag{7}$$

while for $p_{\mathrm{p}} = p_{\mathrm{u}}$ (the choice considered in [5] and other literature we are aware of), we have

$$\mathcal{S} = \log_2 e \left(1 - \frac{\tau}{T} \right) \sum_{k=1}^{K} a_k \tau p_{\mathrm{u}}^2 + \mathcal{O}\left(p_{\mathrm{u}}^3 \right). \tag{8}$$

[1] The achievable ergodic rate for the case of $\beta_{\ell\ell k} = 1$ and $\beta_{\ell i k} = \beta$ $(i \neq \ell)$, for all k, was derived in [5], see Eq. (73).

Interestingly, at low p_{u}, the sum spectral efficiency scales linearly with N [since $a_k = \beta_{\ell\ell k}^2(N-1)$], even though the number of unknown channel parameters increases. We can see that for the case of p_{p} being fixed regardless of p_{u}, at $p_{\mathrm{u}} \ll 1$, the sum spectral efficiency scales as p_{u}. However, for the case of $p_{\mathrm{p}} = p_{\mathrm{u}}$, at $p_{\mathrm{u}} \ll 1$, the sum spectral efficiency scales as p_{u}^2. The reason is that when p_{u} decreases and, hence, p_{p} decreases, the quality of the channel estimate deteriorates, which leads to a "squaring effect" on the sum spectral efficiency [6].

Consider now the bit energy of a system defined as the transmit power expended divided by the sum spectral efficiency:

$$\eta \triangleq \frac{\frac{\tau}{T} p_{\mathrm{p}} + \left(1 - \frac{\tau}{T}\right) p_{\mathrm{u}}}{\mathcal{S}}. \tag{9}$$

If $p_{\mathrm{p}} = p_{\mathrm{u}}$ as in previous works, we have $\eta = \frac{p_{\mathrm{u}}}{\mathcal{S}}$. Then, from (8), when the transmit power is reduced below a certain threshold, the bit energy increases even when we reduce the power (and, hence, reduce the spectral efficiency). As a result, the minimum bit energy is achieved at a non-zero sum spectral efficiency. Evidently, it is inefficient to operate below this sum spectral efficiency. However, we can operate in this regime if we use a large enough transmit power for uplink pilots, and reduce the transmit power of data. This observation is clearly outlined in the next section.

3 Optimal Resource Allocation

Using different powers for the uplink training and data transmission phases improves the system performance, especially in the wideband regime, where the spectral efficiency is conventionally parameterized as an affine function of the energy per bit [9]. Motivated by this observation, we consider a fundamental resource allocation problem, which adjusts the data power, pilot power, and duration of pilot sequences, to maximize the sum spectral efficiency given in (6). Note that, this resource allocation can be implemented at the BS.

Let P be the total transmit energy constraint for each terminal in a coherence interval. Then, we have

$$\tau p_{\mathrm{p}} + (T - \tau) p_{\mathrm{u}} \leq P. \tag{10}$$

When τp_{p} decreases, we can see from (2) that the effect of noise on the channel estimate escalates, and hence the channel estimate degrades. However, under the total energy constraint (10), $(T - \tau)p_{\mathrm{u}}$ will increase, and hence the system performance may improve. Conversely, we could increase the accuracy of the channel estimate by using more power for training. At the same time, we have to reduce the transmit power for the data transmission phase to satisfy (10). Thus, there are optimal values of τ, p_{p}, and p_{u} which maximize the sum spectral efficiency for given P and T.

Once the total transmit energy per coherence interval and the number of terminals are set, one can adjust the duration of pilot sequences and the transmitted powers of pilots and data to maximize the sum spectral efficiency. More precisely,

$$\mathcal{P}_1 : \begin{cases} \max\limits_{p_{\mathrm{u}}, p_{\mathrm{p}}, \tau} & \mathcal{S} \\ \text{s.t.} & \tau p_{\mathrm{p}} + (T-\tau) p_{\mathrm{u}} = P \\ & p_{\mathrm{p}} \geq 0, p_{\mathrm{u}} \geq 0 \\ & K \leq \tau \leq T, (\tau \in \mathbb{N}) \end{cases} \tag{11}$$

where the inequality of the total energy constraint in (10) becomes the equality in (11), due to the fact that for a given τ and p_{p}, $\mathcal{S}$ is an increasing function of p_{u}, and for a given τ and p_{u}, $\mathcal{S}$ is an increasing function of p_{p}. Hence, $\mathcal{S}$ is maximized when $\tau p_{\mathrm{p}} + (T-\tau) p_{\mathrm{u}} = P$.

Proposition 12 *The optimal pilot duration, τ, of $\mathcal{P}_1$ is equal to the number of terminals K.*

Proof: *Let $\left(\tau^*, p_{\mathrm{p}}^*, p_{\mathrm{u}}^*\right)$ be a solution of $\mathcal{P}_1$. Assume that $\tau^* > K$. Next we choose $\bar{\tau} = K$, $\bar{p}_{\mathrm{p}} = \tau^* p_{\mathrm{p}}^* / K$, and $\bar{p}_{\mathrm{u}} = \frac{P - \tau^* p_{\mathrm{p}}^*}{T-K}$. Clearly, this choice of system parameters $(\bar{\tau}, \bar{p}_{\mathrm{p}}, \bar{p}_{\mathrm{u}})$ satisfies the constraints in (11). From (6) and using the fact that $\bar{\tau}\bar{p}_{\mathrm{p}} = \tau^* p_{\mathrm{p}}^*$, we have $\mathcal{S}\left(\bar{\tau}, \bar{p}_{\mathrm{p}}, \bar{p}_{\mathrm{u}}\right) > \mathcal{S}\left(\tau^*, p_{\mathrm{p}}^*, p_{\mathrm{u}}^*\right)$ which contradicts the assumption. Therefore, $\tau^* = K$.* □

From Proposition 12, $\mathcal{P}_1$ is equivalent to the following optimization problem:

$$\mathcal{P}_2 : \begin{cases} \max\limits_{p_{\mathrm{u}}} & \mathcal{S}|_{p_{\mathrm{p}} = P/K - (T/K-1)p_{\mathrm{u}}} \\ \text{s.t.} & 0 \leq p_{\mathrm{u}} \leq \frac{P}{T-K}. \end{cases} \tag{12}$$

We can efficiently solve $\mathcal{P}_2$ based on the following property:

Proposition 13 *The program $\mathcal{P}_2$ is concave.*

Proof: *See Appendix A.* □

To solve the optimization problem $\mathcal{P}_2$, we can use any nonlinear or convex optimization method to get the globally optimal result. Here, we use the FMINCON function in MATLAB's optimization toolbox.

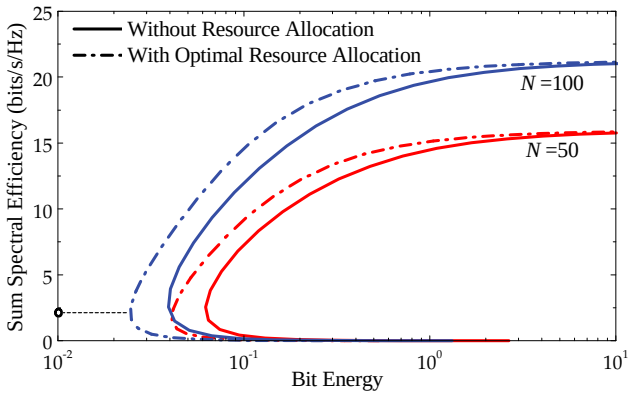


Figure 1: Bit energy versus sum spectral efficiency with and without resource allocation.

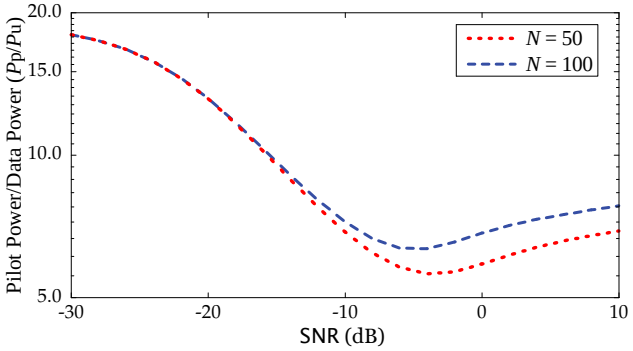


Figure 2: Ratio of the transmit pilot power to the transmit data power.

4 Numerical Results

We consider a cellular network with $L = 7$ hexagonal cells which have a radius of $r_c = 1000$m. Each cell serves 10 terminals ($K = 10$). We choose $T = 200$, corresponding to a coherence bandwidth of 200 KHz and a coherence time of 1 ms. We consider the performance in the cell in the center of the network. We assume that terminals are located uniformly and randomly in each cell and no terminal is closer to the BS than $r_h = 200$m. Large-scale fading is modeled as $\beta_{\ell ik} = z_{\ell ik}/(r_{\ell ik}/r_h)^{\nu}$, where $z_{\ell ik}$ is a log-normal random variable, $r_{\ell ik}$ denotes the distance between the kth terminal in the ith cell and the ℓth BS, and ν is the path loss exponent. We set the standard deviation of $z_{\ell ik}$ to 8dB, and $\nu = 3.8$.

Firstly, we will examine the sum spectral efficiency versus the bit energy obtained from one snapshot generated by the above large-scale fading model. The bit energy is defined in (9). From (9) and (11), we can see that the solution of $\mathcal{P}_1$ also leads to

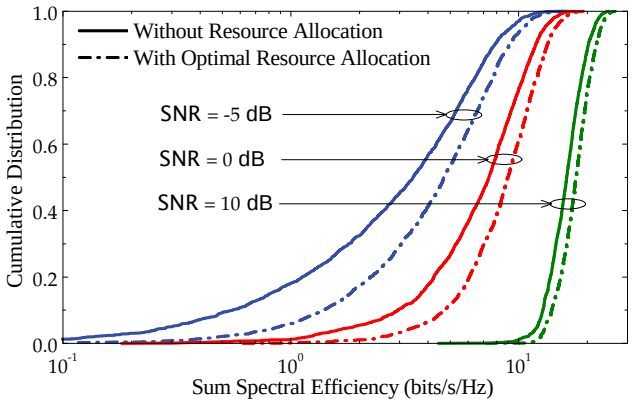


Figure 3: Sum spectral efficiency with and without resource allocation ($N = 100$).

the minimum value of the bit energy. Figure 4 presents the sum spectral efficiency versus the bit energy with optimal resource allocation. As discussed in Section 2.3, the minimum bit energy is achieved at a non-zero spectral efficiency. For example, with optimal resource allocation, at $N = 100$, the minimum bit energy is achieved at a sum spectral efficiency of 2 bits/s/Hz which is marked by a circle in the figure. Below this value, the bit energy increases as the sum spectral efficiency decreases. For a given energy per bit, there are two operating points. Operating below the sum spectral efficiency, at which the minimum energy per bit is obtained, should be avoided.

On a different note, we can see that with optimal resource allocation, the system performance improves significantly. For example, to achieve the same sum spectral efficiency of 10 bits/s/Hz, optimal resource allocation can improve the energy efficiencies by factors of 1.45 and 1.5 compared to the case of no resource allocation with $N = 50$ and $N = 100$, respectively. This dramatic increase underscores the importance of resource allocation in massive MIMO. However, at high bit energy, the squaring effect for the case of no resource allocation disappears and, hence, the advantages of resource allocation diminish. Furthermore, for the same sum spectral efficiency, $\mathcal{S} = 10$ bits/s/Hz, and with resource allocation, by doubling the number of BS antennas from 50 to 100, we can improve the energy efficiency by a factor of 2.2.

The corresponding ratio of the optimal pilot power to the optimal transmitted data power for $N = 50$ and $N = 100$ is shown in Fig. 2. Here, we define $\mathsf{SNR} \triangleq P/T$. Since P is the total transmit energy spent in a coherence interval T and the noise variance is 1, SNR has the interpretation of average transmit SNR and is therefore dimensionless. We can see that at low SNR (or low spectral efficiency), we spend more power during the training phase, and vice versa at high SNR. At low SNR, $p_{\mathrm{p}}/p_{\mathrm{u}} \approx 18$ which leads to $\tau p_{\mathrm{p}}/(T-\tau)p_{\mathrm{u}} \approx 1$. This means that half of the total energy is used for uplink training and the other half is used for data transmission.

Note that the power allocation problem in the low SNR regime is useful since the achievable rate (obtained under the assumption that the estimation error is additive Gaussian noise) is very tight, due to the use of Jensen's bound in [5]. Furthermore, in general, the ratio of the optimal pilot power to the optimal data power does not always monotonically decrease with increasing SNR. We can see from the figure that, when SNR is around -5dB, $p_{\mathrm{p}}/p_{\mathrm{u}}$ increases when SNR increases.

We now consider the cumulative distribution of the sum spectral efficiency obtained from 2000 snapshots of large-scale fading (c.f. Fig. 5). As expected, our resource allocation improves the system performance substantially, especially at low SNR. More importantly, with resource allocation, the sum spectral efficiencies are more concentrated around their means compared to the case of no resource allocation. For example, at SNR $= 0$dB, resource allocation increases the 0.95-likely sum spectral efficiency by a factor of 2 compared to the case of no resource allocation.

5 Conclusion

Conventionally, in massive MIMO, the transmit powers of the pilot signal and data payload signal are assumed to be equal. In this paper, we have posed and answered a basic question about the operation of massive MIMO: How much would the performance improve if the relative energy of the pilot waveform, compared to that of the payload waveform, were chosen optimally? The partitioning of time, or equivalently bandwidth, between pilots and data within a coherence interval was also optimally selected. We found that, with 100 antennas at the BS, by optimally allocating energy to pilots, the energy efficiency can be increased as much as 50%, when each terminal has a throughput of about 1 bit/s/Hz. Typically, when the SNR is low (e.g., around -15dB), at the optimum, the transmit power is then about 10 times higher during the training phase than during the data transmission phase.

Appendix

A Proof of Proposition 13

From (6) and (12), the problem $\mathcal{P}_2$ becomes

$$\mathcal{P}_2 = \begin{cases} \arg\max\limits_{p_{\mathrm{u}}} \ \left(1-\frac{K}{T}\right)\sum_{k=1}^{K}\log_2\left(1+f_k\left(p_{\mathrm{u}}\right)\right) \\ 0 \le p_{\mathrm{u}} \le \frac{P}{T-K} \end{cases} \tag{13}$$

where

$$\begin{aligned} f_k\left(p_{\mathrm{u}}\right) &\triangleq \frac{a_k\left(P-\left(T-K\right)p_{\mathrm{u}}\right)p_{\mathrm{u}}}{b_k\left(P-\left(T-K\right)p_{\mathrm{u}}\right)p_{\mathrm{u}}+c_k p_{\mathrm{u}}+d_k\left(P-\left(T-K\right)p_{\mathrm{u}}\right)+1} \\ &= \frac{a_k}{b_k}-\frac{a_k}{b_k}\frac{c_k p_{\mathrm{u}}+d_k\left(P-\left(T-K\right)p_{\mathrm{u}}\right)+1}{b_k\left(P-\left(T-K\right)p_{\mathrm{u}}\right)p_{\mathrm{u}}+c_k p_{\mathrm{u}}+d_k\left(P-\left(T-K\right)p_{\mathrm{u}}\right)+1}. \end{aligned}$$

The second derivative of $f_k\left(p_{\mathrm{u}}\right)$ can be expressed as follows:

$$\begin{aligned} \omega_k\frac{\partial^2 f_k\left(p_{\mathrm{u}}\right)}{\partial p_{\mathrm{u}}^2} &= -b_k\hat{T}^2(c_k-d_k\hat{T})p_{\mathrm{u}}^3-3b_k\hat{T}^2(d_kP+1)p_{\mathrm{u}}^2 \\ &\quad +3b_k\hat{T}P(d_kP+1)p_{\mathrm{u}}-(d_kP+1)(b_kP^2+c_kP+\hat{T}), \end{aligned} \tag{14}$$

where $\omega_k \triangleq \frac{\left(b_k\left(P-\hat{T}p_{\mathrm{u}}\right)p_{\mathrm{u}}+c_kp_{\mathrm{u}}+d_k\left(P-\hat{T}p_{\mathrm{u}}\right)+1\right)^3}{2a_k}$, and $\hat{T} \triangleq T-K$. Since $P \ge \hat{T}p_{\mathrm{u}}$, we have

$$\begin{aligned} \omega_k\frac{\partial^2 f_k\left(p_{\mathrm{u}}\right)}{\partial p_{\mathrm{u}}^2} &= -b_kc_k\hat{T}^2p_{\mathrm{u}}^3-(d_kP+1)(c_kP+\hat{T}) \\ &\quad -\frac{3}{4}b_k\hat{T}^2p_{\mathrm{u}}^2-b_k\left(P-\frac{3}{2}\hat{T}p_{\mathrm{u}}\right)^2-b_kd_k(P-\hat{T}p_{\mathrm{u}})^3 \le 0. \end{aligned} \tag{15}$$

Since $\omega_k > 0$, $\frac{\partial^2 f_k(p_{\mathrm{u}})}{\partial p_{\mathrm{u}}^2} \le 0$. Therefore, $f_k\left(p_{\mathrm{u}}\right)$ is a concave function in $0 \le p_{\mathrm{u}} \le \frac{P}{T-K}$. Since $\log_2\left(1+x\right)$ is a concave and increasing function, $\log_2\left(1+f_k\left(p_{\mathrm{u}}\right)\right)$ is also a concave function. Finally, using the fact that the summation of concave functions is concave, we conclude the proof of Proposition 13.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[4] K. T. Truong and R. W. Heath Jr., "Effects of channel aging in massive MIMO systems," *IEEE J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[6] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[7] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 5, pp. 357-371, Oct. 2007.

[8] S. Murugesan, E. Uysal-Biyikoglu, and P. Schniter, "Optimization of training and scheduling in the non-coherent SIMO multiple access channel," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1446–1456, Sep. 2007.

[9] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319-1343, June 2002.

Paper H

Large-Scale Multipair Two-Way Relay Networks with Distributed AF Beamforming

Refereed article published in the IEEE Communications Letters 2013.

Large-Scale Multipair Two-Way Relay Networks with Distributed AF Beamforming

Hien Quoc Ngo and Erik G. Larsson

Abstract

We consider a multipair two-way relay network where multiple communication pairs simultaneously exchange information with the help of multiple relay nodes. All nodes are equipped with a single antenna and channel state information (CSI) is available only at the relay nodes. Each relay uses very simple signal processing in a distributed manner, called distributed amplify-and-forward (AF) relaying. *A closed-form expression for the achievable rate is derived. We show that the distributed AF scheme outperforms conventional orthogonal relaying. When the number of relay nodes is large, the distributed AF relaying scheme can achieve the capacity scaling given by the cut-set upper bound. Furthermore, when the number of relay nodes grows large, the transmit powers of each terminal and of the relay can be made inversely proportional to the number of relay nodes while maintaining a given quality-of-service. If the transmit power of each terminal is kept fixed, the transmit power of each relay node can be scaled down inversely proportional to the square of the number of relays.*

1 Introduction

The *multipair one-way relay channel*, where multiple sources simultaneously transmit signals to their destinations through the use of a multiple relay nodes, has attracted substantial interest [1–3]. In [1, 2], the authors proposed a transmission scheme where the beamforming weights at the relays are obtained under the assumption that all relay nodes can cooperate. A simple distributed beamforming scheme that requires only local channel state information (CSI) at the relays, and which performs well with a large number of relay nodes, was proposed in [3]. With one-way protocols, the half-duplex constraint at the relays imposes a pre-log factor 1/2 for the data rate and hence, limits the spectral efficiency. To overcome this spectral efficiency loss in the one-way relay channel, the *multipair two-way relay channel* has recently been considered. Many transmission schemes have been proposed for this channel [4, 5]. However, those studies considered multipair systems where only one relay (equipped with multiple antennas) participates in the transmission. Multiple single-antenna relays supporting multiple communication pairs were considered in [6–8]. In [6], the weighting coefficient at each relay was designed to minimize the transmit power at the relays under a given received signal-to-interference-plus-noise ratio constraint at each terminal. By contrast, the objective function of [7] was the the sum rate. These works assume that there is a central processing center. A distributed beamforming scheme where the relay weighting coefficient is designed at each relay was proposed in [8]. In [8], the relays use zero-forcing to suppress the interpair interference. However, this scheme assumed that each relay has CSI of all relay-terminal pairs. This requires cooperation between the relay nodes for the CSI exchange.

In this paper, we propose and analyze a distributed amplify-and-forward (AF) relaying scheme for *multipair two-way relay channels* which does not require cooperation between the relay nodes. Our scheme is suitable for dense networks where there are many idle nodes willing to act as relays. Here, we assume that the relays have perfect knowledge of local CSI, that is of the channels from each terminal to the relay. The fundamental basis of our proposed scheme is that when the number of relays is large, and under certain assumptions on the channel gains (e.g. zero mean, independent and uniformly bounded variance), the channel vectors between the terminals and relays are pairwisely nearly orthogonal. There is also empirical support for the near-orthogonality assumption, most notably in the large scale MIMO literature [9]. This makes it possible for the relays to use very simple signal processing.

The work that is most closely related to this paper is [3]. In [3] the authors investigated the scaling law of the power efficiency in the multipair one-way relay channel. By contrast, here, we consider the two-way relay channel. We derive a closed-form expression of a lower bound on the capacity which is valid for a finite number of relay nodes. The resulting expression is simple and yields useful insight.

We show that when the number of relays M grows large, the distributed AF relaying achieves the cut-set upper bound on the capacity. Furthermore, when M is large, we achieve the following power scaling laws: (i) the transmit powers of each terminal and of each relay can be scaled $\propto 1/M$ with no performance reduction; and (ii) if the transmit power of each terminal is fixed, the transmit power of each relay can scaled $\propto 1/M^2$.

2 Multipair Two-Way Relay Channel Model

Consider a network in which K communication pairs $(\mathtt{T}_{1,k}, \mathtt{T}_{2,k})$, $k = 1, ..., K$, share the same time-frequency resource. Two terminals $\mathtt{T}_{1,k}$ and $\mathtt{T}_{2,k}$ exchange their information with the help of M relay nodes $\mathtt{R}_m$, $m = 1, 2, ..., M$. Typically, $K \ll M$. All nodes are equipped with a single antenna and use half-duplex operation. We assume that the distance between $\mathtt{T}_{1,k}$ and $\mathtt{T}_{2,k}$ is large or that the link between $\mathtt{T}_{1,k}$ and $\mathtt{T}_{2,k}$ is blocked by obstacles so that there is no direct link between $\mathtt{T}_{1,k}$ and $\mathtt{T}_{2,k}$ that can be exploited. Transmission will take place in both directions (from the terminals to the relays and back) on the same frequency, and we assume that the channels are reciprocal [9].

We further assume that the relay nodes have full CSI, while the terminals have statistical but no instantaneous CSI. The CSI at the relay nodes could be obtained by using training sequences transmitted from the terminals, at a cost of $2K$ symbols per coherence interval. The assumption that the terminals do not have instantaneous CSI is reasonable for practical systems where the number of relay nodes is large. To obtain instantaneous CSI at the terminals, we would have to spend at least M symbols per coherence interval. We show below that due to hardening effects, although the terminals do not have instantaneous CSI, they can near-coherently detect the signals aided by the statistical distribution of the channels.

3 Distributed AF Transmission Scheme

The communication occurs in two phases, as detailed next and in Fig. 1. We assume perfect time synchronization. In [3], the authors have shown that the lack of synchronicity does not have much effect on the system performance.

3.1 Phase I

All terminals simultaneously broadcast their signals to all relay nodes. Let $h_{m,k}$ and $g_{m,k}$ be the channel coefficients from $\mathtt{T}_{1,k}$ to $\mathtt{R}_m$ and from $\mathtt{R}_m$ to $\mathtt{T}_{2,k}$, respectively.

The channel model includes small-scale fading (Rayleigh fading) and large-scale fading, i.e., $h_{m,k} = \sqrt{\alpha_{m,k}}\tilde{h}_{m,k}$ and $g_{m,k} = \sqrt{\beta_{m,k}}\tilde{g}_{m,k}$, where $\tilde{h}_{m,k} \sim \mathcal{CN}(0,1)$, $\tilde{g}_{m,k} \sim \mathcal{CN}(0,1)$. Here, $\alpha_{m,k}$ and $\beta_{m,k}$ represent the large-scale fading. Then, the received signal at $\mathtt{R}_m$ is given by

$$r_m = \sqrt{p_{\mathrm{S}}}\boldsymbol{h}_m^T\boldsymbol{x}_1 + \sqrt{p_{\mathrm{S}}}\boldsymbol{g}_m^T\boldsymbol{x}_2 + w_m, \tag{1}$$

where $\boldsymbol{x}_i \triangleq [x_{i,1} \ ... \ x_{i,K}]^T$, $\sqrt{p_{\mathrm{S}}}x_{i,k}$ is the transmitted signal from $\mathtt{T}_{i,k}$ (the average transmit power of each terminal is p_{S}), $\boldsymbol{h}_m \triangleq [h_{m,1} \ ... \ h_{m,K}]^T$, $\boldsymbol{g}_m \triangleq [g_{m,1} \ ... \ g_{m,K}]^T$, and w_m is AWGN at $\mathtt{R}_m$. We assume that $w_m \sim \mathcal{CN}(0,1)$.

3.2 Phase II — Distributed AF Relaying

All relays broadcast scaled and phase-rotated versions of their received signals to all terminals. The basic idea of distributed AF relaying is as follows. Consider the two-way relay channel with K pairs as a one-way relay channel with $2K$ pairs where the groups of sources and destinations are the same. Then, we apply the relaying scheme for one-way relay channels in [3, Sec. V].[1] We propose to let $\mathtt{R}_m$ transmit the following phase-rotated version of the received signal:

$$x_{\mathtt{R}_m} = \gamma_m \boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^* r_m, \tag{2}$$

where $\boldsymbol{a}_m \triangleq \left[\boldsymbol{h}_m^T \ \boldsymbol{g}_m^T\right]^T$, $\boldsymbol{D} \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_K \\ \boldsymbol{I}_K & \boldsymbol{0} \end{bmatrix}$ is used to permute the signal position to ensure that the signal transmitted from $\mathtt{T}_{1,k}$ arrives at its destination $\mathtt{T}_{2,k}$ and vice versa, and γ_m is a normalization factor which controls the transmit power at $\mathtt{R}_m$, chosen such that $\mathbb{E}\left\{|x_{\mathtt{R}_m}|^2\right\} = p_{\mathrm{R}}$. Hence,[2]

$$\gamma_m = \sqrt{\frac{p_{\mathrm{R}}}{\mathbb{E}\left\{|\boldsymbol{a}_m^H\boldsymbol{D}\boldsymbol{a}_m^*|^2\left(p_{\mathrm{S}}\|\boldsymbol{h}_m\|^2 + p_{\mathrm{S}}\|\boldsymbol{g}_m\|^2 + 1\right)\right\}}} \tag{3}$$

$$= \sqrt{\frac{p_{\mathrm{R}}/4}{p_{\mathrm{S}}\sum_{j=1}^{K}\left(\alpha_{m,j} + \beta_{m,j}\right)\left(\alpha_{m,j}\beta_{m,j} + c_m\right) + c_m}}, \tag{4}$$

where $c_m \triangleq \sum_{i=1}^{K}\alpha_{m,i}\beta_{m,i}$, see Appendix A. Let $n_{2,k}$ be the $\mathcal{CN}(0,1)$ noise at $\mathtt{T}_{2,k}$. Then, the received signal at $\mathtt{T}_{2,k}$ is

$$y_{2,k} = \sum_{m=1}^{M} g_{m,k}x_{\mathtt{R}_m} + n_{2,k}. \tag{5}$$

[1] Considering a multipair one-way relay channel with K sources $\mathtt{T}_{1,k}$, K destinations $\mathtt{T}_{2,k}$, $k = 1, ..., K$, and M relay nodes $\mathtt{R}_m$, $m = 1, ..., M$, the scaled version at $\mathtt{R}_m$ proposed in [3] is $\boldsymbol{g}_m^H\boldsymbol{h}_m^* r_m$.

[2] Note that γ_m could alternatively be chosen depending on the instantaneous CSI which corresponds to a short-term power constraint. However, we choose to use (3) since: i) it yields a tractable form of the achievable rate which enables us to further analyze the system performance; and ii) the law of large numbers guarantees that the denominator of (3) is nearly deterministic unless K is small. Thus, our choice does not affect the obtained insights.

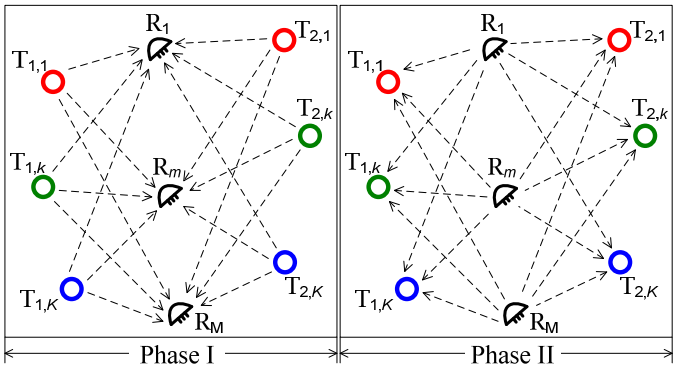


Figure 1: Multipair Two-Way Relaying Network.

When the number of relay nodes M is large, the received signal at $\mathtt{T}_{2,k}$ is dominated by the desired signal part (which includes $x_{1,k}$). As a result, we can obtain noise-free and interference-free communication links when M grows without bound. A more detailed analysis is given in the next section.

3.3 Asymptotic $(M \to \infty, K < \infty)$ Performance

In this section, we provide basic insights into the performance of our proposed scheme when $M \to \infty$ for fixed K and p_{R}. We will show that our proposed scheme performs well when M is large. From (1), (2), and (5), we have

$$y_{2,k} = \underbrace{\sqrt{p_{\mathrm{S}}}\sum_{m=1}^{M} p_{m,k}x_{1,k}}_{\mathcal{L}_1} + \underbrace{\sqrt{p_{\mathrm{S}}}\sum_{j\neq k}^{K}\sum_{m=1}^{M} p_{m,j}x_{1,j} + \sqrt{p_{\mathrm{S}}}\sum_{j=1}^{K}\sum_{m=1}^{M} q_{m,j}x_{2,j}}_{\mathcal{L}_2}$$
$$+ \underbrace{\sum_{m=1}^{M} \gamma_m g_{m,k}\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^* w_m + n_{2,k}}_{\mathcal{L}_3}, \tag{6}$$

where $p_{m,j} \triangleq \gamma_m g_{m,k}\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^* h_{m,j}$ and $q_{m,j} \triangleq \gamma_m g_{m,k}\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^* g_{m,j}$. Here $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ represent the desired signal, multi-terminal interference, and noise effects, respectively. We have

$$\begin{aligned}\mathbb{E}\{p_{m,k}\} &= 2\gamma_m \mathbb{E}\left\{ g_{m,k}\left(\sum_{i=1}^{K} g_{m,i}^* h_{m,i}^*\right) h_{m,k}\right\} \\ &= 2\gamma_m \alpha_{m,k}\beta_{m,k}.\end{aligned} \tag{7}$$

We assume that $\mathtt{Var}\{p_{m,k}\}$, $m = 1, ..., M$, are uniformly bounded, i.e., $\exists c < \infty$: $\mathtt{Var}\{p_{m,k}\} \leq c$, $\forall m, k$ [10]. Since $p_{m,k}$, $m = 1, 2, ..., M$, are independent, it follows from Tchebyshev's theorem [10] that

$$\frac{1}{M}\mathcal{L}_1 - \frac{1}{M}\sqrt{p_{\mathrm{S}}}\sum_{m=1}^{M} 2\gamma_m \alpha_{m,k}\beta_{m,k}x_{1,k} \underset{M\to\infty}{\overset{P}{\to}} 0, \tag{8}$$

where $\overset{P}{\to}$ denotes convergence in probability. Similarly, since $\mathbb{E}\{q_{m,j}\} = 0$ and $\mathbb{E}\left\{\gamma_m g_{m,k}\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^* w_m\right\} = 0$, we have

$$\frac{1}{M}\mathcal{L}_2 \underset{M\to\infty}{\overset{P}{\to}} 0, \ \frac{1}{M}\mathcal{L}_3 \underset{M\to\infty}{\overset{P}{\to}} 0. \tag{9}$$

We can see from (8) and (9) that, when M is large, the power of the desired signal grows as M^2, while the power of the interference and noise grows more slowly. As a result, with an unlimited number of relay nodes, the effects of interference, noise, and fast fading disappear. More precisely, as $M \to \infty$,

$$\frac{y_{2,k}}{M} - \sqrt{p_{\mathrm{S}}}\frac{\sum_{m=1}^{M} 2\gamma_m \alpha_{m,k}\beta_{m,k}}{M}x_{1,k} \overset{P}{\to} 0. \tag{10}$$

The received signal includes only the desired signal and hence, the capacity increases without bound.

4 Achievable Rate for Finite M

In this section, we derive a closed-form expression of the achievable rate for finite M which can be used to draw more precise quantitative conclusions about the performance of the distributed AF relaying scheme. The terminals do not have instantaneous CSI, but they know the statistical distributions of the channels. Hence, the terminals must use the mean of the effective channel gain to coherently detect the desired signals [3].

Consider the link $\mathtt{T}_{1,k} \to$ Relays $\to \mathtt{T}_{2,k}$. From (6), the received signal at $\mathtt{T}_{2,k}$ can be rewritten as the desired signal $\sqrt{p_{\mathrm{S}}}\,\mathbb{E}\left\{\sum_{m=1}^{M} p_{m,k}\right\}x_{1,k}$ plus a remaining term which is considered as the effective noise. This effective noise is uncorrelated with the desired signal. Then, Gaussian noise represents the worst case, and we obtain the following achievable rate:

$$R_{2,k} = \frac{1}{2}\log_2\left(1 + \frac{p_{\mathrm{S}}\left|\mathbb{E}\left\{\sum_{m=1}^{M} p_{m,k}\right\}\right|^2}{p_{\mathrm{S}}\mathtt{Var}\left\{\sum_{m=1}^{M} p_{m,k}\right\} + \mathtt{MT}_k + \mathtt{AN}_k}\right), \tag{11}$$

where the pre-log factor of 1/2 is due to the half-duplex relaying, $\mathtt{Var}\{x\}$ denotes the variance of a RV x, and

$$\mathtt{MT}_k = p_{\mathrm{S}} \sum_{j \neq k}^{K} \mathbb{E}\left\{ \left| \sum_{m=1}^{M} p_{m,j} \right|^2 \right\} + p_{\mathrm{S}} \sum_{j=1}^{K} \mathbb{E}\left\{ \left| \sum_{m=1}^{M} q_{m,j} \right|^2 \right\} \tag{12}$$

$$\mathtt{AN}_k = \sum_{m=1}^{M} \gamma_m^2 \, \mathbb{E}\left\{ \left| g_{m,k} \boldsymbol{a}_m^H \boldsymbol{D} \boldsymbol{a}_m^* \right|^2 \right\} + 1. \tag{13}$$

We now derive a closed-form expression of the achievable rate which is easier to evaluate and to obtain engineering insights from.

Proposition 14 *With distributed AF relaying, the achievable rate of the communication link* $\mathtt{T}_{1,k} \rightarrow$ *Relays* $\rightarrow \mathtt{T}_{2,k}$ *is given by*

$$R_{2,k} = \frac{1}{2} \log_2 \left(1 + \frac{4 p_{\mathrm{S}} \left(\sum_{m=1}^{M} \gamma_m \alpha_{m,k} \beta_{m,k} \right)^2}{p_{\mathrm{R}} \sum\limits_{m=1}^{M} \beta_{m,k} + 4 p_{\mathrm{S}} \sum\limits_{m=1}^{M} \sum\limits_{j=1}^{K} \gamma_m^2 \alpha_{m,k} \beta_{m,k}^2 \left(\alpha_{m,j} + \beta_{m,j} \right) + \varsigma_k + 1} \right), \tag{14}$$

where $\varsigma_k \triangleq 4 p_{\mathrm{S}} \sum\limits_{m=1}^{M} \gamma_m^2 \beta_{m,k}^2 \left(2 \alpha_{m,k} \beta_{m,k} + c_m + \frac{\alpha_{m,k}}{p_{\mathrm{S}}} \right)$.

Proof: *See Appendix B.* □

4.1 Discussion of Results

For simplicity, we next consider a simplified case where the large-scale fading is neglected, i.e., $\alpha_{m,k} = \beta_{m,k} = 1$, for all m, k. The same insights will be straightforwardly obtained for the case when the large-scale fading is taken into account. Substituting $\alpha_{m,k} = \beta_{m,k} = 1$ into (14), we get

$$R_{2,k} = \frac{1}{2} \log_2 \left(1 + \frac{p_{\mathrm{S}} p_{\mathrm{R}} M / K}{p_{\mathrm{S}} p_{\mathrm{R}} \left(2K + 5 + \frac{2}{K} \right) + \frac{p_{\mathrm{R}}}{K} \left(K + 1 \right) + \frac{2 p_{\mathrm{S}}}{M} \left(K + 1 \right) + \frac{1}{M}} \right).$$

We can see that $R_{2,k}$ increases with M, and decreases with K. When the number of relay nodes goes to infinity, $R_{2,k} \rightarrow \infty$. This lower bound on the rate coincides with the asymptotic (but exact) rate obtained in Section 3.3 and hence, the achievable rate (14) is very tight at large M.

4.1.1 Achievability of the Network Capacity

If p_{S} and $E_{\mathrm{R}} = Mp_{\mathrm{R}}$ (total transmit power of all relays) are fixed regardless of M, then $R_{2,k} = \frac{1}{2}\log_2 M + \mathcal{O}(1)$, as $M \to \infty$. This result coincides with the one which is obtained by using the cut-set upper bound on the network capacity of MIMO relay networks where all terminals are equipped with a single antenna [11]. Note that the result obtained in [11] relies on the assumption that the relay and destination nodes have instantaneous CSI. Here, we assume that only the relays have instantaneous CSI. In particular, with our proposed technique, the sum rate scales as $K\log_2 M + \mathcal{O}(1)$ at large M which is identical to the cut-set bound on the sum capacity of our considered multipair two-way relay network.[3]

4.1.2 Power Scaling Laws

(i) If $p_{\mathrm{S}} = E_{\mathrm{S}}/M$ and $p_{\mathrm{R}} = E_{\mathrm{R}}/M$, where E_{S} and E_{R} are fixed regardless of M, then

$$R_{2,k} \to \frac{1}{2}\log_2\left(1 + \frac{E_{\mathrm{S}}E_{\mathrm{R}}}{E_{\mathrm{R}}(K+1)+K}\right), \text{ as } M \to \infty, \tag{15}$$

which implies that when M is large, we can cut the transmit power $p_{\mathrm{S}} \propto 1/M$ without any performance reduction.

(ii) If p_{S} and E_{R} are fixed regardless of M, and $p_{\mathrm{R}} = E_{\mathrm{R}}/M^2$, then

$$R_{2,k} \to \frac{1}{2}\log_2\left(1 + \frac{p_{\mathrm{S}}E_{\mathrm{R}}}{2p_{\mathrm{S}}K(K+1)+K}\right), \text{ as } M \to \infty. \tag{16}$$

We can see that when M is large, the transmit power of each relay node can be reduced proportionally to $1/M^2$ with no performance degradation. As a result, the transmit power of each relay node can be very small.

5 Numerical Results and Discussion

In this section, we examine the sum rate of our proposed scheme. For comparison, we also consider the sum rate of multipair one-way relaying proposed in [3], and

[3] Suppose that all terminals $\mathtt{T}_{1,k}$ can cooperate and all terminals $\mathtt{T}_{2,k}$ can also cooperate. Then we have a two-way relay network with two terminals each equipped with K antennas, and M single-antenna relays. From [12], the network capacity of this resulting system is $K\log_2 M + \mathcal{O}(1)$. Clearly, this resulting system has greater capacity than the original one. Thus, an upper bound on the sum capacity of our multipair network is $K\log_2 M + \mathcal{O}(1)$.

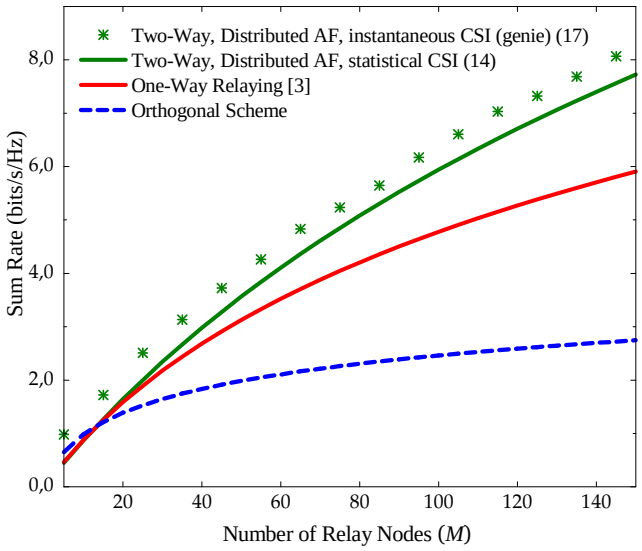


Figure 2: Sum rate versus the number of relay nodes ($K = 5$, $p_{\mathrm{S}} = 10$ dB, and $\alpha_{m,k} = \beta_{m,k} = 1$).

the sum rate of the conventional orthogonal scheme where the transmission of each pair is assigned different time slots or frequency bands. In addition, we consider the sum rate of our scheme but with a genie receiver (instantaneous CSI) at the terminals. For this case, the achievable rate of the link $\mathtt{T}_{1,k} \to$ Relays $\to \mathtt{T}_{2,k}$ is

$$R_{2,k} = \frac{1}{2}\,\mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{S}}\left|\sum_{m=1}^{M} p_{m,k}\right|^2}{\mathtt{MT}_k + \mathtt{AN}_k}\right)\right\}. \tag{17}$$

We choose $K = 5$, $p_{\mathrm{S}} = 10$ dB, and $\alpha_{m,k} = \beta_{m,k} = 1$. We assume that the total transmit powers for the two phases are the same, i.e., $2Kp_{\mathrm{S}} = Mp_{\mathrm{R}}$. Furthermore, for fair comparison, the total transmit powers of all schemes are the same.

Figure 3 shows the sum rate versus the number of relay nodes for the different transmission schemes. We can see that the number of relay nodes has a very strong impact on the performance. The sum rate increases significantly when we increase M. For small M ($\lesssim 10$), owing to inter-terminal interference, our proposed scheme performs worse than the orthogonal scheme. However, when M grows large, the effect of inter-terminal interference and noise dramatically reduces and hence, our proposed scheme outperforms the orthogonal scheme. Compared with the one-way relaying proposed in [3], our distributed AF relaying scheme is better, and the advantage increases when M increases. The gain stems from the reduced pre-log penalty (from 1/2 to 1), however, our scheme suffers from more interference and therefore, the gain is somewhat less than a doubling. When M is small, the inter-terminal interference cannot be notably reduced and hence, our scheme is not better

than the one-way relaying scheme. Furthermore, the performance gap between the cases with instantaneous (genie) and statistical CSI at the terminals is small. This implies that using the mean of the effective channel gain for signal detection is fairly good.

Appendix

A Derivation of (4)

To compute γ_m, we need to compute $\mathbb{E}\left\{|\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^*|^2\right\}$, $\mathbb{E}\left\{|\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^*|^2 \|\boldsymbol{h}_m\|^2\right\}$, and $\mathbb{E}\left\{|\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^*|^2 \|\boldsymbol{g}_m\|^2\right\}$. We have

$$\begin{aligned}\mathbb{E}\left\{\left|\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^*\right|^2 \|\boldsymbol{h}_m\|^2\right\} &= 4\sum_{k=1}^{K}\mathbb{E}\left\{\left|\sum_{i=1}^{K} g_{m,i}^* h_{m,i}^* h_{m,k}\right|^2\right\} \\ &\overset{(a)}{=} 4\sum_{k=1}^{K}\sum_{i=1}^{K}\mathbb{E}\left\{\left|g_{m,i}^* h_{m,i}^* h_{m,k}\right|^2\right\} \\ &\overset{(b)}{=} 8\sum_{k=1}^{K}\alpha_{m,k}^2\beta_{m,k} + 4\sum_{k=1}^{K}\sum_{i\neq k}^{K}\alpha_{m,k}\alpha_{m,i}\beta_{m,i}, \end{aligned} \tag{18}$$

where (a) comes from the fact that $g_{m,i}^* h_{m,i}^* h_{m,k}$, $i = 1, ..., K$, are zero-mean mutual uncorrelated RVs, and (b) follows by using the identity $\mathbb{E}\left\{|x|^2\right\} = \sigma^2$ and $\mathbb{E}\left\{|x|^4\right\} = 2\sigma^4$, where $x \sim \mathcal{CN}\left(0, \sigma^2\right)$. Similarly, we obtain $\mathbb{E}\left\{|\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^*|^2 \|\boldsymbol{g}_m\|^2\right\} = 4\sum_{k=1}^{K}\beta_{m,k}\left(\alpha_{m,k}\beta_{m,k} + c_m\right)$, and $\mathbb{E}\left\{|\boldsymbol{a}_m^H \boldsymbol{D}\boldsymbol{a}_m^*|^2\right\} = 4\sum_{k=1}^{K}\alpha_{m,k}\beta_{m,k}$. Thus, we get (4).

B Proof of Proposition 14

From (11), we need to compute $\texttt{Var}\left\{\sum_{m=1}^{M} p_{m,k}\right\}$, $\texttt{MT}_k$, and $\texttt{AN}_k$. Since $p_{m,k}$, $m = 1, ..., M$, are independent, we have

$$\texttt{Var}\left\{\sum_{m=1}^{M} p_{m,k}\right\} = \sum_{m=1}^{M}\left(\mathbb{E}\left\{|p_{m,k}|^2\right\} - 4\gamma_m^2\alpha_{m,k}^2\beta_{m,k}^2\right). \tag{19}$$

By using the same technique as in Appendix A, we obtain

$$\mathtt{Var}\left\{\sum_{m=1}^{M} p_{m,k}\right\} = 4\sum_{m=1}^{M} \gamma_m^2 \alpha_{m,k}\beta_{m,k}\left(2\alpha_{m,k}\beta_{m,k}+c_m\right). \tag{20}$$

Similarly, we obtain

$$\begin{aligned}\mathtt{MT}_k = {} & 4p_{\mathrm{S}}\sum_{j\neq k}^{K}\sum_{m=1}^{M} \gamma_m^2 \alpha_{m,j}\beta_{m,k}\left(\alpha_{m,k}\beta_{m,k}+\alpha_{m,j}\beta_{m,j}+c_m\right) \\ & + 4p_{\mathrm{S}}\sum_{j=1}^{K}\sum_{m=1}^{M} \gamma_m^2 \beta_{m,j}\beta_{m,k}\left(\alpha_{m,k}\beta_{m,k}+\alpha_{m,j}\beta_{m,j}+c_m\right) \\ & + 4p_{\mathrm{S}}\sum_{m=1}^{M} \gamma_m^2 \beta_{m,k}^2\left(2\alpha_{m,k}\beta_{m,k}+c_m\right), \end{aligned} \tag{21}$$

$$\mathtt{AN}_k = 4\sum_{m=1}^{M} \gamma_m^2 \beta_{m,k}\left(\alpha_{m,k}\beta_{m,k}+c_m\right)+1. \tag{22}$$

Substituting (7), (20), (21), and (22) into (11), we obtain (14).

References

[1] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Gazor, "Multiple peer-to- peer communications using a network of relays," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3053–3062, Aug. 2009.

[2] M. Fadel, A. El-Keyi, and A. Sultan, "QOS-constrained multiuser peer-to-peer amplify-and-forward relay beamforming," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1397–1408, Mar. 2012.

[3] A. F. Dana and B. Hassibi, "On the power efficiency of sensory and ad-hoc wireless networks" *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2890–2914, July 2006.

[4] C. Y. Leow, Z. Ding, K. Leung, and D. Goeckel, "On the study of analogue network coding for multi-pair, bidirectional relay channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 670–681, Feb. 2011.

[5] M. Tao and R.Wang, "Linear precoding for multi-pair two-way MIMO relay systems with max-min fairness," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5361–5370, Oct. 2012.

[6] T. Wang, B. P. Ng, Y. Zhang, and M. H. Er, "Multiple peer-to-peer communications for two-way relay networks," in *Proc. Intern. Conf. Inf., Commun. Signal Process.*, Dec. 2011.

[7] J. Zhang, F. Roemer, and M. Haardt, "Distributed beamforming for two-way relaying networks with individual power constraints," in *Proc. Forty-Sixth Asilomar Conf. Signals, Syst. Comput. (ACSSC)*, Nov. 2012.

[8] C. Wang, H. Chen, Q. Yin, A. Feng, and A. Molisch, "Multi-user two-way relay networks with distributed beamforming," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3460–3471, Oct. 2011.

[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[10] H. Cramér, *Random Variables and Probability Distributions.* Cambridge, UK: Cambridge University Press, 1970.

[11] H. Bölcskei, R. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, June 2006.

[12] R. Vaze and R. W. Heath, Jr., "Capacity scaling for MIMO two-way relaying," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2007.

Paper I

Spectral Efficiency of the Multipair Two-Way Relay Channel with Massive Arrays

Refereed article published in Proc. ACSSC 2013.

Spectral Efficiency of the Multipair Two-Way Relay Channel with Massive Arrays

Hien Quoc Ngo and Erik G. Larsson

Abstract

We consider a multipair two-way relay channel where multiple communication pairs share the same time-frequency resource and a common relay node. We assume that all users have a single antenna, while the relay node is equipped with a very large antenna array. We consider two transmission schemes: (I) separate-training zero-forcing (ZF) and (II) a new proposed coupled-training ZF. For both schemes, the channels are estimated at the relay by using training sequences, assuming time-division duplex operation. The relay processes the received signals using ZF. With the separate-training ZF, the channels from all users are estimated separately. By contrast, with the coupled-training ZF, the relay estimates the sum of the channels from two users of a given communication pair. This reduces the amount of resources spent in the training phase. Self-interference reduction is also proposed for these schemes. When the number of relay antennas grows large, the effects of interpair interference and self-interference can be neglected. The transmit power of each user and of the relay can be made inversely proportional to the square root of the number of relay antennas while maintaining a given quality-of-service. We derive a lower bound on the capacity which enables us to evaluate the spectral efficiency. The coupled-training ZF scheme is preferable for the high-mobility environment, while the separate-training ZF scheme is preferable for the low-mobility environment.

1 Introduction

Two-way relay channels (TWRCs) with multiple-antennas at the participating nodes have been broadly investigated since they can reap all the benefits from two-way and multiple-input multiple-output (MIMO) techniques [1, 2]. With MIMO technology, the TWRC has been extended to *multipair two-way relay channels* where several pairs of users simultaneously communicate with the aid of relays [3]. The main challenge of this system is interpair interference (interference from other communication pairs). A simple method to avoid interpair interference is to let different pairs communicate on orthogonal channels. This is bandwidth inefficient, and higher rates can be achieved if multiple pairs share the same time-frequency resource. Many advanced techniques have been introduced to reduce the effect of interpair interference, such as dirty-paper coding or interference alignment techniques. However, these techniques significantly increase the complexity of the system. Simpler schemes, such as linear processing, are more preferable from a complexity point of view but typically offer somewhat inferior performance.

The MIMO multipair TWRC with linear processing at the relay has been studied in [4–6]. One well known scheme is zero-forcing (ZF), which can eliminate the interpair interference. In [4], the authors consider decode-and-forward protocols where the relay decodes the transmitted signals in the multiple-access phase, and then applies ZF precoding in the broadcast phase. A ZF scheme for amplify-and-forward relaying is proposed in [5]. The papers [4–6] assumed that perfect channel state information (CSI) is available at the relay and at the users. However, in reality, all channels have to be estimated. Therefore, the CSI is not perfectly known which may drastically degrade the system performance. More importantly, in the multipair TWRC, many communication pairs are severed simultaneously and, hence, to obtain CSI, substantial time-frequency resources have to be allocated to the transmission of pilots, which reduces the overall spectral efficiency, especially in high mobility environments.

Very recently, there has been a great deal of interest in large scale (a.k.a. massive) MIMO where the transceivers are equipped with very large antenna arrays (a hundred or more antennas) [7, 8]. With very large arrays, interpair interference can be significantly reduced by using simple processing such as ZF, maximum-ratio combining/transmission (MRC/MRT), even in the presence of poor-quality CSI, due to the asymptotic orthogonality between the channel vectors. Furthermore, the transmit power can be drastically reduced due to the array gain. Relay systems where the relay node is equipped with a very large array were studied in [9]. Reference [9] considered multipair one-way relay channels with simple linear processing at the relay, assuming that the relay and users have perfect CSI. It is shown that by using a very large array at the relay, the transmit power of each user and of the relay can be reduced inversely proportional to the number of relay antennas with no performance degradation.

Inspired by the above discussion, in this paper, we propose and analyze two transmission schemes for multipair TWRCs with a very large antenna array at the relay. These schemes are based on ZF processing, and estimates all channels using pilots, relying on TDD operation and channel reciprocity. The analysis yields lower bounds on the capacity, taking into account the errors and spectral efficiency loss imposed by the pilot transmission and channel estimation. In the first scheme, called *(I) separate-training ZF*, the relay estimates the channels from all users separately and applies ZF precoding. In second scheme, called *(II) coupled-training ZF*, the channels are not estimated individually, instead, the relay estimates the sum of the channels from two users in a given pair. We show that, when the number of relay antennas M is large, we can scale down the transmit powers of each user and of the relay proportionally to $1/\sqrt{M}$ and at the same time increase the spectral efficiency K times by simultaneously serving K pairs of users. In this paper, some derivations and proofs are omitted due to space constraints.

2 System Models and Transmission Schemes

Consider two groups of users $\mathcal{U}_1$ and $\mathcal{U}_2$, where each group includes K single-antenna users. The kth user in group 1, $\mathcal{U}_{1,k}$, wants to exchange information with the kth user in group 2, $\mathcal{U}_{2,k}$, $k = 1, 2, ..., K$, with the help of a common relay, R. The relay R is equipped with M antennas, see Fig. 1. We assume that $M \gg K$. We further assume that all communication links share the same time-frequency resource. The transmission is bidirectional and thus operates in TDD. Within each coherence interval, the information exchange occurs in three phases: the training phase, the payload multiple-access phase, and the payload broadcast phase, see Fig. 2. In Fig. 2, and throughout, T is the length of the coherence interval and τ is the part of the coherence interval used for training.

We next introduce the transmission scheme for our considered system in general. Then, in Sections 2.2.1 and 2.2.2, we specialize the description to the two proposed schemes (I) separate-training ZF and (II) coupled-training ZF.

2.1 General Transmission Scheme

2.1.1 The First Phase — Training

The relay estimates the channels based on pilots transmitted from the users.[1] Let $\boldsymbol{G}_i \in \mathbb{C}^{M \times K}$ be the channel matrix between the relay and the K users in group i,

[1] Alternatively, CSI could first be acquired at each user via transmitted pilots from the relay and then provided to the relay over a reverse (feedback) link. Clearly, since $M \gg K$, this would be very inefficient since the channel estimation overhead will be proportional to M.

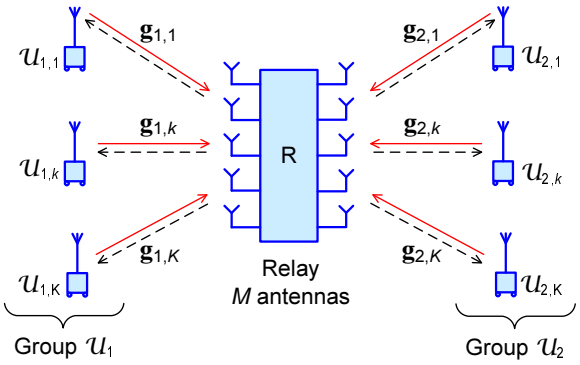


Figure 1: Multipair two-way relay channel with K communication pairs, and an M-antenna relay.

$i = 1, 2$. We assume that $\boldsymbol{G}_i$ has i.i.d. $\mathcal{CN}(0, 1)$ elements. Here, for simplicity of the presentation, we neglect the effects of large-scale fading and path loss.

Each user is assigned a (possibly non-unique in general) pilot sequence of length τ symbols. Let $\sqrt{p_\mathrm{p}}\boldsymbol{\Phi}_i \in \mathbb{C}^{K\times\tau}$ be a matrix whose kth row contains the pilot sequence used by $\mathcal{U}_{i,k}$, $i = 1, 2$, where $p_\mathrm{p} = \tau p_\mathrm{u}$, and p_u is the average transmit power of each user. All users from both groups simultaneously transmit pilot sequences to the relay. The received pilots at the relay can be stacked into a matrix as follows:

$$\boldsymbol{Y}_{\mathrm{R,p}} = \sqrt{p_\mathrm{p}}\boldsymbol{G}_1\boldsymbol{\Phi}_1 + \sqrt{p_\mathrm{p}}\boldsymbol{G}_2\boldsymbol{\Phi}_2 + \boldsymbol{Z}_\mathrm{R}, \tag{1}$$

where $\boldsymbol{Z}_\mathrm{R} \in \mathbb{C}^{M\times\tau}$ is the AWGN at the relay, with i.i.d. $\mathcal{CN}(0, 1)$ elements. The relay will use the above received pilot (1) to estimate the channels. The channel estimate is then used for signal processing at the relay. The design of pilot sequences as well as the channel estimation scheme will be addressed later for each specific transmission scheme.

2.1.2 The Second Phase — Multiple-Access Transmission of Payload Data

All users from both groups simultaneously transmit their data to the relay node. The received signal at the relay node is given by

$$\boldsymbol{y}_\mathrm{R} = \sqrt{p_\mathrm{u}}\boldsymbol{G}_1\boldsymbol{x}_1 + \sqrt{p_\mathrm{u}}\boldsymbol{G}_2\boldsymbol{x}_2 + \boldsymbol{n}_\mathrm{R}, \tag{2}$$

where $\boldsymbol{x}_i \triangleq [x_{i,1} \ldots x_{i,K}]^T$, $i = 1, 2$, $\sqrt{p_\mathrm{u}}x_{i,k}$ is the transmitted signal from the kth user in the ith group (the average transmitted power of each user is p_u); and $\boldsymbol{n}_\mathrm{R} \in \mathbb{C}^M$ is the AWGN vector at the relay, distributed as $\boldsymbol{n}_\mathrm{R} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$.

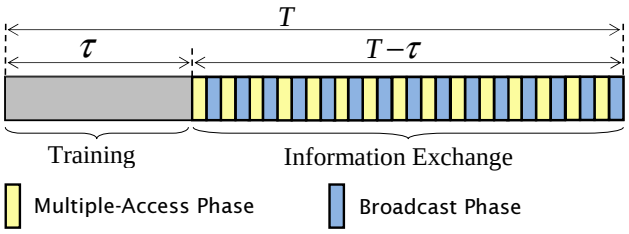


Figure 2: TDD transmission protocol. Here, T is the coherence interval.

2.1.3 The Third Phase — Broadcast of Payload Data

The relay processes the received signal $\boldsymbol{y}_{\mathsf{R}}$ according to a linear processing strategy, and then it broadcasts the processed version to all users. More precisely, the relay sends $\boldsymbol{x}_{\mathsf{R}} = \alpha \boldsymbol{W} \boldsymbol{y}_{\mathsf{R}}$ to all users, where $\boldsymbol{W}$ is a transformation matrix of dimension $M \times M$, and α is a normalization constant. The normalization constant α is chosen to satisfy a long-term total transmit power constraint at the relay, i.e., $\mathsf{Tr}\left(\mathsf{E}\{\boldsymbol{x}_{\mathsf{R}} \boldsymbol{x}_{\mathsf{R}}^H\}\right) = p_{\mathrm{R}}$. Hence,

$$\alpha = \sqrt{\frac{p_{\mathrm{R}}}{\mathsf{Tr}\left(\mathsf{E}\left\{\boldsymbol{W}\left(p_{\mathrm{u}} \boldsymbol{G}_1 \boldsymbol{G}_1^H + p_{\mathrm{u}} \boldsymbol{G}_2 \boldsymbol{G}_2^H + \boldsymbol{I}_M\right) \boldsymbol{W}^H\right\}\right)}}. \tag{3}$$

Equation (3) is obtained under the assumption that $x_{i,k}$, $i = 1, 2$, $k = 1, 2, ..., K$, are independent. This assumption is reasonable since each user independently sends its own data.

The received vector at the K users in group i is given by

$$\boldsymbol{y}_i = \boldsymbol{G}_i^T \boldsymbol{x}_{\mathsf{R}} + \boldsymbol{n}_i = \alpha \boldsymbol{G}_i^T \boldsymbol{W} \boldsymbol{y}_{\mathsf{R}} + \boldsymbol{n}_i, \tag{4}$$

where $\boldsymbol{n}_i$ is the AWGN vector at the K users in group i, $i = 1, 2$, distributed as $\boldsymbol{n}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_K)$.

2.1.4 Self-interference Reduction

After receiving signals, each user reduces self-interference prior to decoding. Self-interference is the interference caused by relaying of the signal that the user transmitted in the multiple-access phase. Without loss of generality, we consider the link $\mathcal{U}_{2,k} \to \mathsf{R} \to \mathcal{U}_{1,k}$. Let $\boldsymbol{g}_{i,k}$ be the kth column of $\boldsymbol{G}_i$. Then, from (4), the received signal at user $\mathcal{U}_{1,k}$ is

$$\begin{aligned} y_{1,k} = {} & \alpha \sqrt{p_{\mathrm{u}}} \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,k} x_{2,k} + \alpha \sqrt{p_{\mathrm{u}}} \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{1,k} x_{1,k} \\ & + \alpha \sqrt{p_{\mathrm{u}}} \sum_{i \neq k}^{K} \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{1,i} x_{1,i} + \alpha \sqrt{p_{\mathrm{u}}} \sum_{i \neq k}^{K} \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,i} x_{2,i} + \alpha \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{n}_R + n_{1,k}, \end{aligned} \tag{5}$$

where $n_{1,k}$ is the kth element of $\boldsymbol{n}_1$. The second term of (5) includes user $\mathcal{U}_{1,k}$'s own transmitted signal, $x_{1,k}$, so it represents self-interference. If user $\mathcal{U}_{1,k}$ had full CSI, it could completely remove the self-interference by subtracting $\alpha\sqrt{p_{\mathrm{u}}}\boldsymbol{g}_{1,k}^T\boldsymbol{W}\boldsymbol{g}_{1,k}x_{1,k}$ from $y_{1,k}$. However, only approximate CSI is available at the relay from the channel estimation in the training phase. So $\mathcal{U}_{1,k}$ cannot remove its self-interference. But when M grows large, we have that

$$\boldsymbol{g}_{1,k}^T\boldsymbol{W}\boldsymbol{g}_{1,k} \to \beta, \text{ as } M \to \infty, \tag{6}$$

for some deterministic constant β. The value of β will be provided later for each specific transmission scheme. By using this fact, we propose the following simple self-interference reduction scheme: subtract $\alpha\sqrt{p_{\mathrm{u}}}\beta x_{1,k}$ from $y_{1,k}$ prior to decoding. From (6), as M grows large, the self-interference can be significantly reduced. The received signal at user $\mathcal{U}_{1,k}$ after using our proposed self-interference reduction scheme is

$$\begin{aligned}\tilde{y}_{1,k} = &\underbrace{\alpha\sqrt{p_{\mathrm{u}}}\boldsymbol{g}_{1,k}^T\boldsymbol{W}\boldsymbol{g}_{2,k}x_{2,k}}_{\text{desired signal}} + \underbrace{\alpha\sqrt{p_{\mathrm{u}}}\left(\boldsymbol{g}_{1,k}^T\boldsymbol{W}\boldsymbol{g}_{1,k} - \beta\right)x_{1,k}}_{\text{self-interference}} \\ &+ \underbrace{\alpha\sqrt{p_{\mathrm{u}}}\sum_{i\neq k}^{K}\boldsymbol{g}_{1,k}^T\boldsymbol{W}\boldsymbol{g}_{1,i}x_{1,i} + \alpha\sqrt{p_{\mathrm{u}}}\sum_{i\neq k}^{K}\boldsymbol{g}_{1,k}^T\boldsymbol{W}\boldsymbol{g}_{2,i}x_{2,i}}_{\text{interpair interference}} + \underbrace{\alpha\boldsymbol{g}_{1,k}^T\boldsymbol{W}\boldsymbol{n}_R + n_{1,k}}_{\text{noise}}.\end{aligned} \tag{7}$$

Specific transmission schemes are proposed in the next section.

2.2 Specific Transmission Schemes

2.2.1 Transmission Scheme I — Separate-Training ZF

We consider separate-training ZF scheme. The relay treats the channel estimate as the true channel and then applies ZF-based receive and transmit beamforming as in [5]. This requires that $M \geq 2K$.[2] For this scheme, in the training phase, the channels $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are estimated separately. This means that the pilot sequences assigned for all users in both groups are pairwisely orthogonal. More precisely, the pilot matrices $\sqrt{p_{\mathrm{p}}}\boldsymbol{\Phi}_i$, $i = 1, 2$, have to satisfy $\boldsymbol{\Phi}_i\boldsymbol{\Phi}_j^H = \delta_{ij}\boldsymbol{I}_K$, where $\delta_{ij} = 1$ when $i = j$ and 0 otherwise. This requires that $\tau \geq 2K$. From (1), the least-squares (LS) channel estimates of $\boldsymbol{G}_i$, $i = 1, 2$ are

$$\hat{\boldsymbol{G}}_i = \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{Y}_{\mathsf{R},\mathrm{p}}\boldsymbol{\Phi}_i^H = \boldsymbol{G}_i + \frac{1}{\sqrt{p_{\mathrm{p}}}}\tilde{\boldsymbol{Z}}_i, \tag{8}$$

[2] Note that, in [5], the authors assumed that the relay and users have perfect CSI. Only if CSI is perfect, this technique can completely remove the interference. However, in practice, CSI always has to be estimated.

where $\tilde{\boldsymbol{Z}}_i \triangleq \boldsymbol{Z}_{\mathsf{R}}\boldsymbol{\Phi}_i^H$, $i = 1, 2$. Since $\boldsymbol{\Phi}_i\boldsymbol{\Phi}_i^H = \boldsymbol{I}_K$, $\tilde{\boldsymbol{Z}}_i$ has i.i.d. $\mathcal{CN}(0, 1)$ elements.

In the third (broadcast) phase, the relay treats the channel estimates as the true channels, and performs ZF as in [5]. Therefore, the transformation matrix $\boldsymbol{W}$ is given by

$$\boldsymbol{W} = \boldsymbol{W}_{\text{SZF}} \triangleq \hat{\boldsymbol{G}}^* \left(\hat{\boldsymbol{G}}^T \hat{\boldsymbol{G}}^*\right)^{-1} \boldsymbol{D} \left(\hat{\boldsymbol{G}}^H \hat{\boldsymbol{G}}\right)^{-1} \hat{\boldsymbol{G}}^H, \tag{9}$$

where $\hat{\boldsymbol{G}} \triangleq \left[\hat{\boldsymbol{G}}_1\ \hat{\boldsymbol{G}}_2\right]$, and $\boldsymbol{D} \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_K \\ \boldsymbol{I}_K & \boldsymbol{0} \end{bmatrix}$.

Note that, due to the channel estimation error, interference cannot be removed completely as in [5]. We next determine the value of the constant β in (6) for this transmission scheme. By using the law of large numbers (LLN), we have

$$\boldsymbol{g}_{1,k}^T \boldsymbol{W}_{\text{SZF}} \boldsymbol{g}_{1,k} \overset{a.s.}{\to} 0, \text{ as } M \to \infty, \tag{10}$$

where $\overset{a.s.}{\to}$ denotes almost sure convergence. Therefore, $\beta = 0$. This means that, when $M \to \infty$, self-interference due to the channel estimate error is automatically canceled out, hence, no subtraction of self-interference as in (7) is needed in this case.

2.2.2 Transmission Scheme II — Coupled-Training ZF

The drawback of scheme I (separate-training ZF) is that $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are estimated separately and, thus, we have to spend at least $2K$ symbols for training. We now propose a scheme in which the relay does not estimate $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ separately. Instead, it estimates $\boldsymbol{G}_1 + \boldsymbol{G}_2$. This conserves resources: operationally, $\mathcal{U}_{1,k}$ and $\mathcal{U}_{2,k}$ use the exact same resources for channel estimation. Therefore, we need to spend only K symbols on training in each coherence interval. The intuition behind this idea is that when the number of relay antennas is large, the columns of $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are asymptotically orthogonal and, hence, the system performance will not be affected much when we use $\boldsymbol{G}_1 + \boldsymbol{G}_2$ instead of $\boldsymbol{G}_1$ or $\boldsymbol{G}_2$ for the relay processing.

With this scheme, in the training phase, K users in group 1 are assigned orthogonal pilot sequences, and the same set of orthogonal pilot sequences is reused in group 2. More precisely, $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2 = \boldsymbol{\Phi}$ which satisfies $\boldsymbol{\Phi}\boldsymbol{\Phi}^H = \boldsymbol{I}_K$. From (1), the LS channel estimate of $\boldsymbol{G}_1 + \boldsymbol{G}_2$ is given by

$$\hat{\boldsymbol{G}}_{\text{s}} = \frac{1}{\sqrt{p_{\text{p}}}} \boldsymbol{Y}_{\mathsf{R},\text{p}} \boldsymbol{\Phi}^H = \boldsymbol{G}_1 + \boldsymbol{G}_2 + \frac{1}{\sqrt{p_{\text{p}}}} \tilde{\boldsymbol{Z}}_{\mathsf{R}}, \tag{11}$$

where $\tilde{\boldsymbol{Z}}_{\mathsf{R}} \triangleq \boldsymbol{Z}_{\mathsf{R}}\boldsymbol{\Phi}^H$ has i.i.d. $\mathcal{CN}(0, 1)$ elements.

We next use a transformation matrix at the relay which is based on the ZF principle. In the first step, the relay uses ZF to combine the signals transmitted from users, and then in the second step, it uses ZF precoding to forward data to all users. The transformation matrix is

$$\boldsymbol{W} = \boldsymbol{W}_{\mathtt{CZF}} \triangleq \hat{\boldsymbol{G}}_{\mathrm{s}}^{*} \left(\hat{\boldsymbol{G}}_{\mathrm{s}}^{T} \hat{\boldsymbol{G}}_{\mathrm{s}}^{*} \right)^{-1} \left(\hat{\boldsymbol{G}}_{\mathrm{s}}^{H} \hat{\boldsymbol{G}}_{\mathrm{s}} \right)^{-1} \hat{\boldsymbol{G}}_{\mathrm{s}}^{H}. \tag{12}$$

We next determine β in (6) for self-interference reduction. By using the LLN, we have

$$\boldsymbol{g}_{1,k}^{T} \boldsymbol{W}_{\mathtt{CZF}} \boldsymbol{g}_{1,k} \overset{a.s.}{\to} (2 + 1/p_{\mathrm{p}})^{-2}, \text{ as } M \to \infty. \tag{13}$$

Therefore, $\beta = (2 + 1/p_{\mathrm{p}})^{-2}$, hence the subtraction in (7) will take effect.

3 Asymptotic $M \to \infty$ Analysis

In this section, we provide basic insights into the performance of our proposed schemes with an unlimited number of relay antennas. We consider the received signal at user $\mathcal{U}_{1,k}$ prior to decoding as (7). By using the LLN and Lindeberg-Lévy central limit theorem, we obtain the following results.

Proposition 15 *Assume that the transmit powers of each user, p_{u}, and of the relay node, p_{R}, are fixed. Then, we have*

$$\tilde{y}_{1,k}/\sqrt{M} \overset{a.s.}{\to} \sqrt{\alpha_0} x_{2,k}, \text{ as } M \to \infty, \tag{14}$$

where $\tilde{y}_{1,k}$ is given by (7), and

$$\alpha_0 = \begin{cases} \frac{p_{\mathrm{R}}}{2K(1+1/p_{\mathrm{p}})} & \textit{for the separate-training ZF scheme (I)} \\ \frac{p_{\mathrm{R}}}{2K(2+1/p_{\mathrm{p}})} & \textit{for the coupled-training ZF scheme (II).} \end{cases}$$

Equation (14) implies that, when $M \to \infty$, the effects of channel estimation error, self-interference, interpair interference, and noise disappear. Our proposed schemes perform very well at large M. The received signal includes only the desired signal and, hence, the capacity increases without bound as $M \to \infty$.

Proposition 16 *Assume that the transmit powers of each user, p_{u}, and of the relay node, p_{R}, are $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$ and $p_{\mathrm{R}} = E_{\mathrm{R}}/\sqrt{M}$, where E_{u} and E_{R} are fixed regardless of M. Then, as $M \to \infty$, we have*

$$\tilde{y}_{1,k} \overset{d}{\to} \alpha_1 \sqrt{\tau} E_{\mathrm{u}} x_{2,k} + \alpha_1 \tilde{n}_{\mathrm{R},k} + n_{1,k}, \tag{15}$$

where $\xrightarrow{d}$ *denotes convergence in distribution, and*

$$\alpha_1 = \begin{cases} \sqrt{\frac{\tau E_{\mathrm{u}} E_{\mathrm{R}}}{2K(\tau E_{\mathrm{u}}^2+1)}} & \textit{for the separate-training ZF scheme (I)} \\ \sqrt{\frac{\tau E_{\mathrm{u}} E_{\mathrm{R}}}{K(2\tau E_{\mathrm{u}}^2+1)}} & \textit{for the coupled-training ZF scheme (II)} \end{cases}$$

and where $\tilde{n}_{\mathsf{R},k}$ *is a* $\mathcal{CN}(0,1)$ *random variable.*

We can see that when $M \to \infty$, the effects of interference and channel fading disappear. The channel is equivalent to a deterministic Gaussian channel independently of M. This implies that, by using large antenna arrays at the relay, the transmit powers of each user and of the relay can be scaled down $\propto 1/\sqrt{M}$ and maintain a nonzero asymptotic capacity:

$$C_{1,k}^{\infty} = \log_2\left(1 + \tau\alpha_1^2 {E_{\mathrm{u}}}^2 / \left(\alpha_1^2 + \tau E_{\mathrm{u}}\right)\right). \tag{16}$$

4 Lower Bound on the Capacity for Finite M

In this section, we derive a lower bound on the capacity which can be used to draw more precise quantitative conclusions on our proposed transmission schemes for finite M. This lower bound is obtained by using the technique of [10], assuming that the interference plus noise has a Gaussian distribution. Consider the communication link $\mathcal{U}_{2,k} \to \mathsf{R} \to \mathcal{U}_{1,k}$. The received signal at $\mathcal{U}_{1,k}$ is given by (7). With our proposed transmission schemes, the relay has knowledge of the channel estimates, while the users do not have this information. However, it is reasonable to assume that user $\mathcal{U}_{1,k}$ does know the statistical properties of the channel. From (7), we have

$$\tilde{y}_{1,k} = \underbrace{\alpha\sqrt{p_{\mathrm{u}}}\mathtt{E}\left\{\boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,k}\right\} x_{2,k}}_{\text{desired signal}} + \underbrace{\tilde{n}_{1,k}}_{\text{effective noise}}, \tag{17}$$

where $\tilde{n}_{1,k}$ is considered as an effective noise term, given by

$$\begin{aligned} \tilde{n}_{1,k} \triangleq\ & \alpha\sqrt{p_{\mathrm{u}}}\left(\boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,k} - \mathtt{E}\left\{\boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,k}\right\}\right) x_{2,k} \\ & + \alpha\sqrt{p_{\mathrm{u}}}\left(\boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{1,k} - \beta\right) x_{1,k} + \alpha\sqrt{p_{\mathrm{u}}} \sum_{i \neq k}^{K} \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{1,i} x_{1,i} \\ & + \alpha\sqrt{p_{\mathrm{u}}} \sum_{i \neq k}^{K} \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,i} x_{2,i} + \alpha \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{n}_R + n_{1,k}. \end{aligned} \tag{18}$$

We can easily show that the "desired signal" and the "effective noise" $\tilde{n}_{1,k}$ in (17) are uncorrelated. Therefore, the channel (17) is equivalent to a deterministic gain channel with uncorrelated additive noise. By using the fact that the worst-case

uncorrelated additive noise is independent Gaussian noise of same variance, we obtain the following lower on the capacity of the communication link $\mathcal{U}_{2,k} \to \mathsf{R} \to \mathcal{U}_{1,k}$:

$$R_{1,k} = \log_2 \left(1 + \frac{\alpha^2 p_{\mathrm{u}} \left| \mathtt{E} \left\{ \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,k} \right\} \right|^2}{\alpha^2 p_{\mathrm{u}} \mathbb{V}\mathrm{ar} \left(\boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,k} \right) + \mathtt{SI}_k + \mathtt{IP}_k + \mathtt{AN}_k} \right), \tag{19}$$

where $\mathtt{SI}_k$, $\mathtt{IP}_k$, and $\mathtt{AN}_k$ represent the self-interference, interpair interference, and additive noise effects, respectively:

$$\mathtt{SI}_k \triangleq \alpha^2 p_{\mathrm{u}} \mathtt{E} \left\{ \left| \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{1,k} - \beta \right|^2 \right\}, \tag{20}$$

$$\mathtt{IP}_k \triangleq \alpha^2 p_{\mathrm{u}} \sum_{i \neq k}^{K} \mathtt{E} \left\{ \left| \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{1,i} \right|^2 + \left| \boldsymbol{g}_{1,k}^T \boldsymbol{W} \boldsymbol{g}_{2,i} \right|^2 \right\}, \tag{21}$$

$$\mathtt{AN}_k \triangleq \alpha^2 \mathtt{E} \left\{ \left\| \boldsymbol{g}_{1,k}^T \boldsymbol{W} \right\|^2 \right\} + 1. \tag{22}$$

Remark 12 *If $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$ and $p_{\mathrm{R}} = E_{\mathrm{R}}/\sqrt{M}$ where E_{u} and E_{R} are fixed regardless of M, then as $M \to \infty$, $R_{1,k}$ converges to the asymptotic capacity given by* (16)*. Hence, the bound is very tight at large M.*

5 Numerical Results

In this section, we examine the spectral efficiency of our proposed schemes. The spectral efficiency is defined as the sum-rate (in bits) per channel use. During a coherence interval of T symbols, we spend τ symbols for training, and the remaining interval is used for the payload data exchange. Therefore, the spectral efficiency is given by

$$\mathcal{S}_{\mathtt{b}} = \frac{T - \tau}{2T} \sum_{k=1}^{K} \left(R_{1,k} + R_{2,k} \right),$$

where $R_{1,k}$ is given by (19), and $R_{2,k}$ is the lower bound on the capacity of the link $\mathcal{U}_{1,k} \to \mathsf{R} \to \mathcal{U}_{2,k}$ obtained by interchanging 1 and 2 in (19)–(22).

As a baseline for comparison, we also consider an *orthogonal scheme* that employs orthogonal channel access to avoid interference. With this scheme, at least $\tau = 2K$ training symbols are used to estimate $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ separately. The remaining duration (i.e. $T - \tau$) is divided into $4K$ parts in which each transmission (from $\mathcal{U}_{1,k}$ to R, or from R to $\mathcal{U}_{2,k}$, or the reverse directions) is performed. At the relay, MRC/MRT

is employed as it maximizes the effective SNR. Therefore, the spectral efficiency is given by

$$\mathcal{S}_{\mathsf{0}} = \frac{T-\tau}{4KT}\sum_{k=1}^{K}\left(R_{1,k}^{\mathsf{0}} + R_{2,k}^{\mathsf{0}}\right),$$

where

$$R_{1,k}^{\mathsf{0}} = R_{2,k}^{\mathsf{0}} = \log_2\left(1+\frac{c_{\mathsf{0}}^2 p_{\mathrm{u}}\left|\mathtt{E}\left\{\boldsymbol{g}_{1,k}^T\boldsymbol{W}_k^{\mathsf{0}}\boldsymbol{g}_{2,k}\right\}\right|^2}{c_{\mathsf{0}}^2 p_{\mathrm{u}}\mathbb{V}\mathrm{ar}\left(\boldsymbol{g}_{1,k}^T\boldsymbol{W}_k^{\mathsf{0}}\boldsymbol{g}_{2,k}\right) + c_{\mathsf{0}}^2\mathtt{E}\left\{\left\|\boldsymbol{g}_{1,k}^T\boldsymbol{W}_k^{\mathsf{0}}\right\|^2\right\} + 1}\right),$$

and where $\boldsymbol{W}_k^{\mathsf{0}} \triangleq \hat{\boldsymbol{g}}_{1,k}^*\hat{\boldsymbol{g}}_{2,k}^H$, $\hat{\boldsymbol{g}}_{i,k}$ is the kth column of $\hat{\boldsymbol{G}}_i$, and

$$c_{\mathsf{0}} \triangleq \sqrt{\frac{p_{\mathrm{R}}}{\mathsf{Tr}\left(\mathtt{E}\left\{\hat{\boldsymbol{g}}_{1,k}^*\hat{\boldsymbol{g}}_{2,k}^H\left(p_{\mathrm{u}}\boldsymbol{g}_{2,k}\boldsymbol{g}_{2,k}^H + \boldsymbol{I}_M\right)\hat{\boldsymbol{g}}_{2,k}\hat{\boldsymbol{g}}_{1,k}^T\right\}\right)}}.$$

In all examples, we choose $K = 20$, $\tau = 2K$ for the separate-training ZF and orthogonal schemes, while $\tau = K$ for the coupled-training ZF scheme. Furthermore, for our proposed schemes, we assume that the total transmit powers during the multiple-access phase and the broadcast phase are the same, i.e., $p_{\mathrm{R}} = 2Kp_{\mathrm{u}}$. For our proposed schemes, we choose $p_{\mathrm{u}} = 0$ dB, while for the orthogonal scheme we must have $p_{\mathrm{u}} = 40T/(T+\tau)$ to guarantee that the total transmitted energies in one symbol interval for all schemes are the same.

Fig. 3 shows the spectral efficiency versus the number of relay antennas for the different schemes, with $T = 50$. We can see that our proposed transmission schemes outperform the orthogonal scheme, especially for large M. For small M, the separate-training ZF scheme is the best. However when M grows large, the coupled-training ZF scheme is better. This is due to the fact that, the spectral efficiency is affected by the pre-log factor and SINR. Compared with the separate-training ZF scheme, the coupled-training ZF scheme has larger pre-log factor (since it uses less training symbols to estimate the channels), but it has lower SINR since it surfers from larger interference. When the number of relay antennas is large, interpair interference and self-interference can be notably reduced; as a consequence, the pre-log factor has a larger impact on the system performance.

Next, we consider the effect of the coherence interval length T on the system performance for different transmission schemes. Figure 4 shows the spectral efficiency versus the length of the coherence interval for $M = 500$. Again, our proposed schemes outperform the orthogonal scheme for all T. For short coherence intervals, the coupled scheme performs better than the separate-training ZF scheme and vice versa at large T. The reason is that, at large T, the impact of the training duration can be ignored and, hence, the pre-log factors for both schemes are the same, while the separate-training ZF scheme has an advantage of larger SINR.

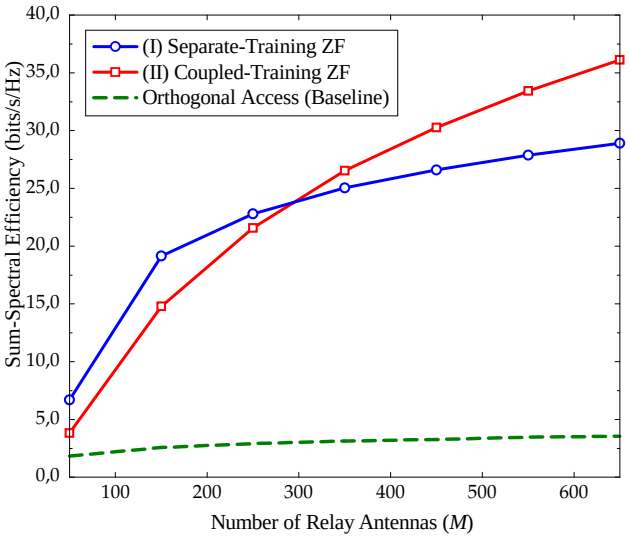


Figure 3: Spectral efficiency versus the number of relay antennas for different transmission schemes ($K = 20$ and $T = 50$).

6 Conclusion

We considered the multipair TWRC where the relay is equipped with a very large antenna array. We proposed two specific transmission schemes: (I) separate-training ZF and (II) coupled-training ZF, both based on ZF-precoding. In the first scheme, the relay estimates the channels from all users separately. In the second scheme, the relay estimates the sum of channels from two users of a given communication pair. We provided closed-form lower bounds on the sum-capacity of the two schemes, taking into account the effects of channel estimation.

Depending on the operating regime, scheme I is better than scheme II and vice versa. Generally, in a high-mobility environment, the coupled-training scheme performs better than the separate-training scheme and vice versa. The gain of our proposed schemes over conventional orthogonal access is significant.

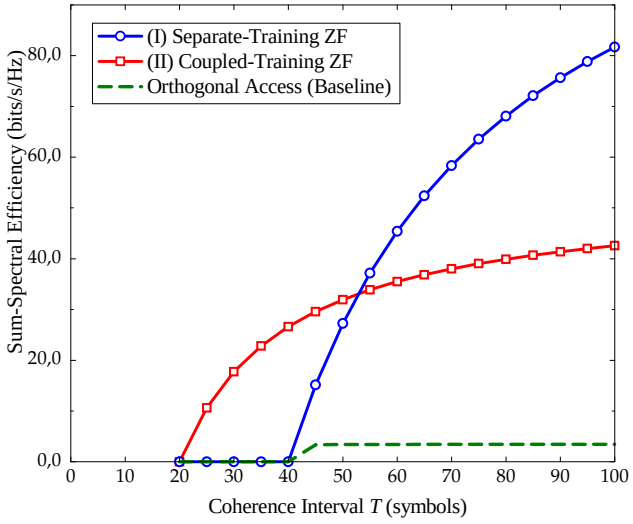


Figure 4: Spectral efficiency versus the length of the coherence interval for different transmission schemes ($K = 20$ and $M = 500$).

References

[1] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, 2009.

[2] R. F. Wyrembelski, T. J. Oechtering, and H. Boche, "MIMO Gaussian bidirectional broadcast channels with common messages," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2950–2959, Sep. 2011.

[3] J. Joung and A. H. Sayed, "Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1833–1846, 2010.

[4] C. Esli and A. Wittneben, "Multiuser MIMO two-way relaying for cellular communications," in *Proc. IEEE 19th Int. Symp. Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Sep. 2008.

[5] E. Yilmaz, R. Zakhour, D. Gesbert, and R. Knopp, "Multi-pair two-way relay channel with multiple antenna relay station," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2010.

[6] M. Tao and R. Wang, "Linear precoding for multi-pair two-way MIMO relay systems with max-min fairness," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5361–5370, Oct. 2012.

[7] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[8] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[9] H. A. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen, and E. G. Larsson, "Multi-pair amplify-and-forward relaying with very large antenna arrays," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2013.

[10] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. Asilomar Conf. Signals, Systems, Comput.*, Oct. 2006.

Paper J

Multipair Full-Duplex Relaying with Massive Arrays and Linear Processing

Refereed article published in the IEEE Journal on
Selected Areas in Communications 2014.

Multipair Full-Duplex Relaying with Massive Arrays and Linear Processing

Hien Quoc Ngo, Himal A. Suraweera, Michail Matthaiou, and Erik G. Larsson

Abstract

We consider a multipair decode-and-forward relay channel, where multiple sources transmit simultaneously their signals to multiple destinations with the help of a full-duplex relay station. We assume that the relay station is equipped with massive arrays, while all sources and destinations have a single antenna. The relay station uses channel estimates obtained from received pilots and zero-forcing (ZF) or maximum-ratio combining/maximum-ratio transmission (MRC/MRT) to process the signals. To reduce significantly the loop interference effect, we propose two techniques: i) using a massive receive antenna array; or ii) using a massive transmit antenna array together with very low transmit power at the relay station. We derive an exact achievable rate expression in closed-form for MRC/MRT processing and an analytical approximation of the achievable rate for ZF processing. This approximation is very tight, especially for a large number of relay station antennas. These closed-form expressions enable us to determine the regions where the full-duplex mode outperforms the half-duplex mode, as well as, to design an optimal power allocation scheme. This optimal power allocation scheme aims to maximize the energy efficiency for a given sum spectral efficiency and under peak power constraints at the relay station and sources. Numerical results verify the effectiveness of the optimal power allocation scheme. Furthermore, we show that, by doubling the number of transmit/receive antennas at the relay station, the transmit power of each source and of the relay station can be reduced by $1.5dB$ if the pilot power is equal to the signal power, and by $3dB$ if the pilot power is kept fixed, while maintaining a given quality-of-service.

1 Introduction

Multiple-input multiple-output (MIMO) systems that use antenna arrays with a few hundred antennas for multiuser operation (popularly called "Massive MIMO") is an emerging technology that can deliver all the attractive benefits of traditional MIMO, but at a much larger scale [1–3]. Such systems can reduce substantially the effects of noise, fast fading and interference and provide increased throughput. Importantly, these attractive features of massive MIMO can be reaped using simple signal processing techniques and at a reduction of the total transmit power. Not surprisingly, massive MIMO combined with cooperative relaying is a strong candidate for the development of future energy-efficient cellular networks [3, 4].

On a parallel avenue, full-duplex relaying has received a lot of research interest, for its ability to recover the bandwidth loss induced by conventional half-duplex relaying. With full-duplex relaying, the relay node receives and transmits simultaneously on the same channel [5,6]. As such, full-duplex utilizes the spectrum resources more efficiently. Over the recent years, rapid progress has been made on both theory and experimental hardware platforms to make full-duplex wireless communication an efficient practical solution [7–12]. The benefit of improved spectral efficiency in the full-duplex mode comes at the price of loop interference due to signal leakage from the relay's output to the input [8, 9]. A large amplitude difference between the loop interference and the received signal coming from the source can exceed the dynamic range of the analog-to-digital converter at the receiver side, and, thus, its mitigation is crucial for full-duplex operation [12, 13]. Note that how to overcome the detrimental effects of loop interference is a highly active area in full-duplex research.

Traditionally, loop interference suppression is performed in the antenna domain using a variety of passive techniques that electromagnetically shield the transmit antenna from the receive antenna. As an example, directional antennas can be used to place a null at the receive antenna. Since the distance between the transmit and receive arrays is short, such techniques require significant levels of loop interference mitigation and, hence, are hard to realize. On the other hand, active time domain loop interference cancellation techniques use the knowledge of the interfering signal to pre-cancel the loop interference in the radio frequency signal and achieve higher levels of loop interference suppression. However, they require advanced noise cancellation methods and sophisticated electronic implementation [7]. Yet, MIMO processing provides an effective means of suppressing the loop interference in the spatial domain. With multiple transmit or receive antennas at the full-duplex relay, precoding solutions, such as zero-forcing (ZF), can be deployed to mitigate the loop interference effects. Although sub-optimal in general, a simple ZF-based precoder can completely cancel the loop interference and remove the closed-loop between the relay's input and output. Several papers have considered spatial loop interference suppression; for example, [9] proposes to direct the loop interference of a full-duplex decode-and-forward (DF) relay to the least harmful spatial dimensions.

In [7], assuming a multiple antenna relay, a range of spatial suppression techniques, including precoding and antenna selection, is analyzed. In [14], several antenna sub-set selection schemes are proposed aiming to suppress loop interference at the relay's transmit side. More recently, [15] analyzed several antenna selection schemes for spatial loop interference suppression in a MIMO relay channel.

Different from the majority of existing works in the literature, which consider systems that deploy only few antennas, in this paper we consider a massive MIMO full-duplex relay architecture. The large number of spatial dimensions available in a massive MIMO system can be effectively used to suppress the loop interference in the spatial domain. We assume that a group of K sources communicate with a group of K destinations through a massive MIMO full-duplex relay station. Specifically, in this multipair massive MIMO relay system, we deploy two processing schemes, namely, ZF and maximum ratio combining (MRC)/maximal ratio transmission (MRT) with full-duplex relay operation. Recall that linear processing techniques, such as ZF or MRC/MRT processing, are low-complexity solutions that are anticipated to be utilized in massive MIMO topologies. Their main advantage is that in the large-antenna limit, they can perform as well as non-linear schemes (e.g., maximum-likelihood) [1, 4, 16]. Our system setup could be applied in cellular networks, where several users transmit simultaneously signals to several other users with the help of a relay station (infrastructure-based relaying). Note that, newly evolving wireless standards, such as LTE-Advanced, promote the use of relays (with unique cell ID and right for radio resource management) to serve as low power base stations [17, 18].

We investigate the achievable rate and power efficiency of the aforementioned full-duplex system setup. Moreover, we compare full-duplex and half-duplex modes and show the benefit of choosing one over the other (depending on the loop interference level of the full-duplex mode). Although the current work uses techniques related to those in Massive MIMO, we investigate a substantially different setup. Specifically, previous works related to Massive MIMO systems [1–3, 21] considered the uplink or the downlink of multiuser MIMO channels. In contrast, we consider multipair full-duplex relaying channels with massive arrays at the relay station. As a result, our new contributions are very different from the existing works on Massive MIMO. The main contributions of this paper are summarized as follows:

1. We show that the loop interference can be significantly reduced, if the relay station is equipped with a large receive antenna array or/and is equipped with a large transmit antenna array. At the same time, the inter-pair interference and noise effects disappear. Furthermore, when the number of relay station transmit antennas, N_{tx}, and the number of relay station receive antennas, N_{rx}, are large, we can scale down the transmit powers of each source and of the relay proportionally to $1/N_{\mathrm{rx}}$ and $1/N_{\mathrm{tx}}$, respectively, if the pilot power is kept fixed, and proportionally to $1/\sqrt{N_{\mathrm{rx}}}$ and $1/\sqrt{N_{\mathrm{tx}}}$, respectively, if the pilot power and the data power are the same.

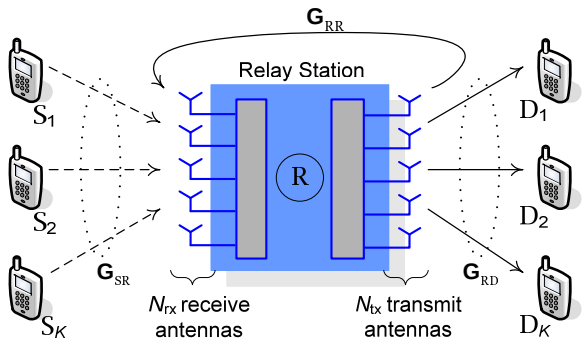


Figure 1: Multipair full-duplex relaying system.

2. We derive exact and approximate closed-form expressions for the end-to-end (e2e) achievable rates of MRC/MRT and ZF processing, respectively. These simple closed-form expressions enable us to obtain important insights as well as to compare full-duplex and half-duplex operation and demonstrate which mode yields better performance. As a general remark, the full-duplex mode improves significantly the overall system performance when the loop interference level is low. In addition, we propose the use of a hybrid mode for each large-scale fading realization, which switches between the full-duplex and half-duplex modes, to maximize the sum spectral efficiency.

3. We design an optimal power allocation algorithm for the data transmission phase, which maximizes the energy efficiency for a desired sum spectral efficiency and under peak power constraints at the relay station and sources. This optimization problem can be approximately solved via a sequence of geometric programs (GPs). Our numerical results indicate that the proposed power allocation improves notably the performance compared to uniform power allocation.

Notation: We use boldface upper- and lower-case letters to denote matrices and column vectors, respectively. The superscripts $()^*$, $()^T$, and $()^H$ stand for the conjugate, transpose, and conjugate-transpose, respectively. The Euclidean norm, the trace, the expectation, and the variance operators are denoted by $\|\cdot\|$, $\mathrm{tr}(\cdot)$, $\mathbb{E}\{\cdot\}$, and $\mathbb{V}\mathrm{ar}(\cdot)$, respectively. The notation $\xrightarrow{a.s.}$ means almost sure convergence, while $\xrightarrow{d}$ means convergence in distribution. Finally, we use $\boldsymbol{z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ to denote a circularly symmetric complex Gaussian vector $\boldsymbol{z}$ with zero mean and covariance matrix $\boldsymbol{\Sigma}$.

2 System Model

Figure 1 shows the considered multipair DF relaying system where K communication pairs $(\mathtt{S}_k, \mathtt{D}_k)$, $k = 1, \ldots, K$, share the same time-frequency resource and a common relay station, $\mathtt{R}$. The kth source, $\mathtt{S}_k$, communicates with the kth destination, $\mathtt{D}_k$, via the relay station, which operates in a full-duplex mode. All source and destination nodes are equipped with a single antenna, while the relay station is equipped with N_{rx} receive antennas and N_{tx} transmit antennas. The total number of antennas at the relay station is $N = N_{\mathrm{rx}} + N_{\mathrm{tx}}$. We assume that the hardware chain calibration is perfect so that the channel from the relay station to the destination is reciprocal [3]. Further, the direct links among $\mathtt{S}_k$ and $\mathtt{D}_k$ do not exist due to large path loss and heavy shadowing. Our network configuration is of practical interest, for example, in a cellular setup, where inter-user communication is realized with the help of a base station equipped with massive arrays.

At time instant i, all K sources $\mathtt{S}_k$, $k = 1, ..., K$, transmit simultaneously their signals, $\sqrt{p_{\mathrm{S}}} x_k[i]$, to the relay station, while the relay station broadcasts $\sqrt{p_{\mathrm{R}}}\boldsymbol{s}[i] \in \mathbb{C}^{N_{\mathrm{tx}} \times 1}$ to all K destinations. Here, we assume that $\mathbb{E}\left\{|x_k[i]|^2\right\} = 1$ and $\mathbb{E}\left\{\|\boldsymbol{s}[i]\|^2\right\} = 1$ so that p_{S} and p_{R} are the average transmit powers of each source and of the relay station. Since the relay station receives and transmits at the same frequency, the received signal at the relay station is interfered by its own transmitted signal, $\boldsymbol{s}[i]$. This is called *loop interference*. Denote by $\boldsymbol{x}[i] \triangleq [x_1[i]\ x_2[i]\ ...\ x_K[i]]^T$. The received signals at the relay station and the K destinations are given by [7]

$$\boldsymbol{y}_{\mathtt{R}}[i] = \sqrt{p_{\mathrm{S}}}\boldsymbol{G}_{\mathtt{SR}}\boldsymbol{x}[i] + \sqrt{p_{\mathrm{R}}}\boldsymbol{G}_{\mathtt{RR}}\boldsymbol{s}[i] + \boldsymbol{n}_{\mathtt{R}}[i], \tag{1}$$

$$\boldsymbol{y}_{\mathtt{D}}[i] = \sqrt{p_{\mathrm{R}}}\boldsymbol{G}_{\mathtt{RD}}^T\boldsymbol{s}[i] + \boldsymbol{n}_{\mathtt{D}}[i], \tag{2}$$

respectively, where $\boldsymbol{G}_{\mathtt{SR}} \in \mathbb{C}^{N_{\mathrm{rx}} \times K}$ and $\boldsymbol{G}_{\mathtt{RD}}^T \in \mathbb{C}^{K \times N_{\mathrm{tx}}}$ are the channel matrices from the K sources to the relay station's receive antenna array and from the relay station's transmit antenna array to the K destinations, respectively. The channel matrices account for both small-scale fading and large-scale fading. More precisely, $\boldsymbol{G}_{\mathtt{SR}}$ and $\boldsymbol{G}_{\mathtt{RD}}$ can be expressed as $\boldsymbol{G}_{\mathtt{SR}} = \boldsymbol{H}_{\mathtt{SR}}\boldsymbol{D}_{\mathtt{SR}}^{1/2}$ and $\boldsymbol{G}_{\mathtt{RD}} = \boldsymbol{H}_{\mathtt{RD}}\boldsymbol{D}_{\mathtt{RD}}^{1/2}$, where the small-scale fading matrices $\boldsymbol{H}_{\mathtt{SR}}$ and $\boldsymbol{H}_{\mathtt{RD}}$ have independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ elements, while $\boldsymbol{D}_{\mathtt{SR}}$ and $\boldsymbol{D}_{\mathtt{RD}}$ are the large-scale fading diagonal matrices whose kth diagonal elements are denoted by $\beta_{\mathtt{SR},k}$ and $\beta_{\mathtt{RD},k}$, respectively. The above channel models rely on the favorable propagation assumption, which assumes that the channels from the relay station to different sources and destinations are independent [3]. The validity of this assumption was demonstrated in practice, even for massive arrays [19]. Also in (1), $\boldsymbol{G}_{\mathtt{RR}} \in \mathbb{C}^{N_{\mathrm{rx}} \times N_{\mathrm{tx}}}$ is the channel matrix between the transmit and receive arrays which represents the loop interference. We model the loop interference channel via the Rayleigh fading distribution, under the assumptions that any line-of-sight component is efficiently reduced by antenna

isolation and the major effect comes from scattering. Note that if hardware loop interference cancellation is applied, $\boldsymbol{G}_{\mathtt{RR}}$ represents the residual interference due to imperfect loop interference cancellation. The residual interfering link is also modeled as a Rayleigh fading channel, which is a common assumption made in the existing literature [7]. Therefore, the elements of $\boldsymbol{G}_{\mathtt{RR}}$ can be modeled as i.i.d. $\mathcal{CN}\left(0,\sigma_{\mathtt{LI}}^2\right)$ random variables, where $\sigma_{\mathtt{LI}}^2$ can be understood as the level of loop interference, which depends on the distance between the transmit and receive antenna arrays or/and the capability of a hardware loop interference cancellation technique [8]. Here, we assume that the distance between the transmit array and the receive array is much larger than the inter-element distance, such that the channels between the transmit and receive antennas are i.i.d.;[1] also, $\boldsymbol{n}_{\mathtt{R}}[i]$ and $\boldsymbol{n}_{\mathtt{D}}[i]$ are additive white Gaussian noise (AWGN) vectors at the relay station and the K destinations, respectively. The elements of $\boldsymbol{n}_{\mathtt{R}}[i]$ and $\boldsymbol{n}_{\mathtt{D}}[i]$ are assumed to be i.i.d. $\mathcal{CN}(0,1)$.

2.1 Channel Estimation

In practice, the channels $\boldsymbol{G}_{\mathtt{SR}}$ and $\boldsymbol{G}_{\mathtt{RD}}$ have to be estimated at the relay station. The standard way of doing this is to utilize pilots [1]. To this end, a part of the coherence interval is used for channel estimation. All sources and destinations transmit simultaneously their pilot sequences of τ symbols to the relay station. The received pilot matrices at the relay receive and transmit antenna arrays are given by

$$\boldsymbol{Y}_{\mathrm{rp}} = \sqrt{\tau p_{\mathrm{p}}}\boldsymbol{G}_{\mathtt{SR}}\boldsymbol{\Phi}_{\mathtt{S}} + \sqrt{\tau p_{\mathrm{p}}}\bar{\boldsymbol{G}}_{\mathtt{RD}}\boldsymbol{\Phi}_{\mathtt{D}} + \boldsymbol{N}_{\mathrm{rp}}, \tag{3}$$

$$\boldsymbol{Y}_{\mathrm{tp}} = \sqrt{\tau p_{\mathrm{p}}}\bar{\boldsymbol{G}}_{\mathtt{SR}}\boldsymbol{\Phi}_{\mathtt{S}} + \sqrt{\tau p_{\mathrm{p}}}\boldsymbol{G}_{\mathtt{RD}}\boldsymbol{\Phi}_{\mathtt{D}} + \boldsymbol{N}_{\mathrm{tp}}, \tag{4}$$

respectively, where $\bar{\boldsymbol{G}}_{\mathtt{SR}} \in \mathbb{C}^{N_{\mathrm{tx}}\times K}$ and $\bar{\boldsymbol{G}}_{\mathtt{RD}} \in \mathbb{C}^{N_{\mathrm{rx}}\times K}$ are the channel matrices from the K sources to the relay station's transmit antenna array and from the K destinations to the relay station's receive antenna array, respectively; p_{p} is the transmit power of each pilot symbol, $\boldsymbol{N}_{\mathrm{rp}}$ and $\boldsymbol{N}_{\mathrm{tp}}$ are AWGN matrices which include i.i.d. $\mathcal{CN}(0,1)$ elements, while the kth rows of $\boldsymbol{\Phi}_{\mathtt{S}} \in \mathbb{C}^{K\times\tau}$ and $\boldsymbol{\Phi}_{\mathtt{D}} \in \mathbb{C}^{K\times\tau}$ are the pilot sequences transmitted from $\mathtt{S}_k$ and $\mathtt{D}_k$, respectively. All pilot sequences are assumed to be pairwisely orthogonal, i.e., $\boldsymbol{\Phi}_{\mathtt{S}}\boldsymbol{\Phi}_{\mathtt{S}}^H = \boldsymbol{I}_K$, $\boldsymbol{\Phi}_{\mathtt{D}}\boldsymbol{\Phi}_{\mathtt{D}}^H = \boldsymbol{I}_K$, and $\boldsymbol{\Phi}_{\mathtt{S}}\boldsymbol{\Phi}_{\mathtt{D}}^H = \boldsymbol{0}_K$. This requires that $\tau \geq 2K$.

We assume that the relay station uses minimum mean-square-error (MMSE) estimation to estimate $\boldsymbol{G}_{\mathtt{SR}}$ and $\boldsymbol{G}_{\mathtt{RD}}$. The MMSE channel estimates of $\boldsymbol{G}_{\mathtt{SR}}$ and $\boldsymbol{G}_{\mathtt{RD}}$ are

[1] For example, consider two transmit and receive arrays which are located on the two sides of a building with a distance of 3m. Assume that the system is operating at 2.6GHz. Then, to guarantee uncorrelation between the antennas, the distance between adjacent antennas is about 6cm, which is half a wavelength. Clearly, 3m $\gg$ 6cm. In addition, if each array is a cylindrical array with 128 antennas, the physical size of each array is about 28cm $\times$29cm [19] which is still relatively small compared to the distance between the two arrays.

given by [20]

$$\hat{\boldsymbol{G}}_{\mathtt{SR}} = \frac{1}{\sqrt{\tau p_{\mathrm{p}}}} \boldsymbol{Y}_{\mathrm{rp}} \boldsymbol{\Phi}_{\mathtt{S}}^{H} \tilde{\boldsymbol{D}}_{\mathtt{SR}} = \boldsymbol{G}_{\mathtt{SR}} \tilde{\boldsymbol{D}}_{\mathtt{SR}} + \frac{1}{\sqrt{\tau p_{\mathrm{p}}}} \boldsymbol{N}_{\mathtt{S}} \tilde{\boldsymbol{D}}_{\mathtt{SR}}, \tag{5}$$

$$\hat{\boldsymbol{G}}_{\mathtt{RD}} = \frac{1}{\sqrt{\tau p_{\mathrm{p}}}} \boldsymbol{Y}_{\mathrm{tp}} \boldsymbol{\Phi}_{\mathtt{D}}^{H} \tilde{\boldsymbol{D}}_{\mathtt{RD}} = \boldsymbol{G}_{\mathtt{RD}} \tilde{\boldsymbol{D}}_{\mathtt{RD}} + \frac{1}{\sqrt{\tau p_{\mathrm{p}}}} \boldsymbol{N}_{\mathtt{D}} \tilde{\boldsymbol{D}}_{\mathtt{RD}}, \tag{6}$$

respectively, where $\tilde{\boldsymbol{D}}_{\mathtt{SR}} \triangleq \left(\frac{\boldsymbol{D}_{\mathtt{SR}}^{-1}}{\tau p_{\mathrm{p}}} + \boldsymbol{I}_K \right)^{-1}$, $\tilde{\boldsymbol{D}}_{\mathtt{RD}} \triangleq \left(\frac{\boldsymbol{D}_{\mathtt{RD}}^{-1}}{\tau p_{\mathrm{p}}} + \boldsymbol{I}_K \right)^{-1}$, $\boldsymbol{N}_{\mathtt{S}} \triangleq \boldsymbol{N}_{\mathrm{rp}} \boldsymbol{\Phi}_{\mathtt{S}}^{H}$ and $\boldsymbol{N}_{\mathtt{D}} \triangleq \boldsymbol{N}_{\mathrm{tp}} \boldsymbol{\Phi}_{\mathtt{D}}^{H}$. Since the rows of $\boldsymbol{\Phi}_{\mathtt{S}}$ and $\boldsymbol{\Phi}_{\mathtt{D}}$ are pairwisely orthogonal, the elements of $\boldsymbol{N}_{\mathtt{S}}$ and $\boldsymbol{N}_{\mathtt{D}}$ are i.i.d. $\mathcal{CN}(0,1)$ random variables. Let $\boldsymbol{\mathcal{E}}_{\mathtt{SR}}$ and $\boldsymbol{\mathcal{E}}_{\mathtt{RD}}$ be the estimation error matrices of $\boldsymbol{G}_{\mathtt{SR}}$ and $\boldsymbol{G}_{\mathtt{RD}}$, respectively. Then,

$$\boldsymbol{G}_{\mathtt{SR}} = \hat{\boldsymbol{G}}_{\mathtt{SR}} + \boldsymbol{\mathcal{E}}_{\mathtt{SR}}, \tag{7}$$

$$\boldsymbol{G}_{\mathtt{RD}} = \hat{\boldsymbol{G}}_{\mathtt{RD}} + \boldsymbol{\mathcal{E}}_{\mathtt{RD}}. \tag{8}$$

From the property of MMSE channel estimation, $\hat{\boldsymbol{G}}_{\mathtt{SR}}$, $\boldsymbol{\mathcal{E}}_{\mathtt{SR}}$, $\hat{\boldsymbol{G}}_{\mathtt{RD}}$, and $\boldsymbol{\mathcal{E}}_{\mathtt{RD}}$ are independent [20]. Furthermore, we have that the rows of $\hat{\boldsymbol{G}}_{\mathtt{SR}}$, $\boldsymbol{\mathcal{E}}_{\mathtt{SR}}$, $\hat{\boldsymbol{G}}_{\mathtt{RD}}$, and $\boldsymbol{\mathcal{E}}_{\mathtt{RD}}$ are mutually independent and distributed as $\mathcal{CN}\left(\boldsymbol{0}, \hat{\boldsymbol{D}}_{\mathtt{SR}}\right)$, $\mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{D}_{\mathtt{SR}} - \hat{\boldsymbol{D}}_{\mathtt{SR}}\right)$, $\mathcal{CN}\left(\boldsymbol{0}, \hat{\boldsymbol{D}}_{\mathtt{RD}}\right)$, and $\mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{D}_{\mathtt{RD}} - \hat{\boldsymbol{D}}_{\mathtt{RD}}\right)$, respectively, where $\hat{\boldsymbol{D}}_{\mathtt{SR}}$ and $\hat{\boldsymbol{D}}_{\mathtt{RD}}$ are diagonal matrices whose kth diagonal elements are $\sigma_{\mathtt{SR},k}^2 \triangleq \frac{\tau p_{\mathrm{p}} \beta_{\mathtt{SR},k}^2}{\tau p_{\mathrm{p}} \beta_{\mathtt{SR},k} + 1}$ and $\sigma_{\mathtt{RD},k}^2 \triangleq \frac{\tau p_{\mathrm{p}} \beta_{\mathtt{RD},k}^2}{\tau p_{\mathrm{p}} \beta_{\mathtt{RD},k} + 1}$, respectively.

2.2 Data Transmission

The relay station considers the channel estimates as the true channels and employs linear processing. More precisely, the relay station uses a linear receiver to decode the signals transmitted from the K sources. Simultaneously, it uses a linear precoding scheme to forward the signals to the K destinations.

2.2.1 Linear Receiver

With the linear receiver, the received signal $\boldsymbol{y}_{\mathtt{R}}[i]$ is separated into K streams by multiplying it with a linear receiver matrix $\boldsymbol{W}^T$ (which is a function of the channel estimates) as follows:

$$\boldsymbol{r}[i] = \boldsymbol{W}^T \boldsymbol{y}_{\mathtt{R}}[i] = \sqrt{p_{\mathrm{S}}} \boldsymbol{W}^T \boldsymbol{G}_{\mathtt{SR}} \boldsymbol{x}[i] + \sqrt{p_{\mathrm{R}}} \boldsymbol{W}^T \boldsymbol{G}_{\mathtt{RR}} \boldsymbol{s}[i] + \boldsymbol{W}^T \boldsymbol{n}_{\mathtt{R}}[i]. \tag{9}$$

Then, the kth stream (kth element of $\boldsymbol{r}[i]$) is used to decode the signal transmitted from $\mathtt{S}_k$. The kth element of $\boldsymbol{r}[i]$ can be expressed as

$$r_k[i] = \underbrace{\sqrt{p_{\mathrm{S}}} \boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} x_k[i]}_{\text{desired signal}} + \underbrace{\sqrt{p_{\mathrm{S}}} \sum_{j \neq k}^{K} \boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},j} x_j[i]}_{\text{interpair interference}} + \underbrace{\sqrt{p_{\mathrm{R}}} \boldsymbol{w}_k^T \boldsymbol{G}_{\mathtt{RR}} \boldsymbol{s}[i]}_{\text{loop interference}} + \underbrace{\boldsymbol{w}_k^T \boldsymbol{n}_{\mathtt{R}}[i]}_{\text{noise}}, \tag{10}$$

where $\boldsymbol{g}_{\mathtt{SR},k}$, $\boldsymbol{w}_k$ are the kth columns of $\boldsymbol{G}_{\mathtt{SR}}$, $\boldsymbol{W}$, respectively, and $x_k[i]$ is the kth element of $\boldsymbol{x}[i]$.

2.2.2 Linear Precoding

After detecting the signals transmitted from the K sources, the relay station uses linear precoding to process these signals before broadcasting them to all K destinations. Owing to the processing delay [7], the transmit vector $\boldsymbol{s}[i]$ is a precoded version of $\boldsymbol{x}[i-d]$, where d is the processing delay. More precisely,

$$\boldsymbol{s}[i] = \boldsymbol{A}\boldsymbol{x}[i-d], \tag{11}$$

where $\boldsymbol{A} \in \mathbb{C}^{N_{\mathtt{tx}} \times K}$ is a linear precoding matrix which is a function of the channel estimates. We assume that the processing delay $d \geq 1$ which guarantees that, for a given time instant, the receive and transmit signals at the relay station are uncorrelated. This is a common assumption for full-duplex systems in the existing literature [8,10].

From (2) and (11), the received signal at $\mathtt{D}_k$ can be expressed as

$$y_{\mathtt{D},k}[i] = \sqrt{p_{\mathrm{R}}}\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_k x_k[i-d] + \sqrt{p_{\mathrm{R}}}\sum_{j\neq k}^{K}\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_j x_j[i-d] + n_{\mathtt{D},k}[i], \tag{12}$$

where $\boldsymbol{g}_{\mathtt{RD},k}$, $\boldsymbol{a}_k$ are the kth columns of $\boldsymbol{G}_{\mathtt{RD}}$, $\boldsymbol{A}$, respectively, and $n_{\mathtt{D},k}[i]$ is the kth element of $\boldsymbol{n}_{\mathtt{D}}[i]$.

2.3 ZF and MRC/MRT Processing

In this work, we consider two common linear processing techniques: ZF and MRC/MRT processing.

2.3.1 ZF Processing

In this case, the relay station uses the ZF receiver and ZF precoding to process the signals. Due to the fact that all communication pairs share the same time-frequency resource, the transmission of a given pair will be impaired by the transmissions of other pairs. This effect is called "interpair interference". More explicitly, for the transmission from $\mathtt{S}_k$ to the relay station, the interpair interference is represented by the term $\sqrt{p_{\mathrm{S}}}\sum_{j\neq k}^{K}\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},j}x_j[i]$, while for the transmission from the relay station to $\mathtt{D}_k$, the interpair interference is $\sqrt{p_{\mathrm{R}}}\sum_{j\neq k}^{K}\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_j x_j[i-d]$. With

ZF processing, interpair interference is nulled out by projecting each stream onto the orthogonal complement of the interpair interference. This can be done if the relay station has perfect channel state information (CSI). However, in practice, the relay station knows only the estimates of CSI. Therefore, interpair interference and loop interference still exist. We assume that $N_{\mathrm{rx}}, N_{\mathrm{tx}} > K$.

The ZF receiver and ZF precoding matrices are respectively given by [21,22]

$$\boldsymbol{W}^T = \boldsymbol{W}_{\mathtt{ZF}}^T \triangleq \left(\hat{\boldsymbol{G}}_{\mathtt{SR}}^H \hat{\boldsymbol{G}}_{\mathtt{SR}}\right)^{-1} \hat{\boldsymbol{G}}_{\mathtt{SR}}^H, \tag{13}$$

$$\boldsymbol{A} = \boldsymbol{A}_{\mathtt{ZF}} \triangleq \alpha_{\mathtt{ZF}} \hat{\boldsymbol{G}}_{\mathtt{RD}}^* \left(\hat{\boldsymbol{G}}_{\mathtt{RD}}^T \hat{\boldsymbol{G}}_{\mathtt{RD}}^*\right)^{-1}, \tag{14}$$

where $\alpha_{\mathtt{ZF}}$ is a normalization constant, chosen to satisfy a long-term total transmit power constraint at the relay, i.e., $\mathbb{E}\left\{\left\|\boldsymbol{s}\left[i\right]\right\|^2\right\} = 1$. Therefore, we have [22]

$$\alpha_{\mathtt{ZF}} \triangleq \sqrt{\frac{N_{\mathrm{tx}} - K}{\sum_{k=1}^{K} \sigma_{\mathtt{RD},k}^{-2}}}. \tag{15}$$

2.3.2 MRC/MRT Processing

The ZF processing neglects the effect of noise and, hence, it works poorly when the signal-to-noise ratio (SNR) is low. By contrast, the MRC/MRT processing aims to maximize the received SNR, by neglecting the interpair interference effect. Thus, MRC/MRT processing works well at low SNRs, and works poorly at high SNRs. With MRC/MRT processing, the relay station uses MRC to detect the signals transmitted from the K sources. Then, it uses the MRT technique to transmit signals towards the K destinations. The MRC receiver and MRT precoding matrices are respectively given by [21,22]

$$\boldsymbol{W}^T = \boldsymbol{W}_{\mathtt{MRC}}^T \triangleq \hat{\boldsymbol{G}}_{\mathtt{SR}}^H, \tag{16}$$

$$\boldsymbol{A} = \boldsymbol{A}_{\mathtt{MRT}} \triangleq \alpha_{\mathtt{MRT}} \hat{\boldsymbol{G}}_{\mathtt{RD}}^*, \tag{17}$$

where the normalization constant $\alpha_{\mathtt{MRT}}$ is chosen to satisfy a long-term total transmit power constraint at the relay, i.e., $\mathbb{E}\left\{\left\|\boldsymbol{s}\left[i\right]\right\|^2\right\} = 1$, and we have [22]

$$\alpha_{\mathtt{MRT}} \triangleq \sqrt{\frac{1}{N_{\mathrm{tx}} \sum_{k=1}^{K} \sigma_{\mathtt{RD},k}^2}}. \tag{18}$$

3 Loop Interference Cancellation with Large Antenna Arrays

In this section, we consider the potential of using massive MIMO technology to cancel the loop interference due to the full-duplex operation at the relay station. Some interesting insights are also presented.

3.1 Using a Large Receive Antenna Array ($N_{\mathrm{rx}} \to \infty$)

The loop interference can be canceled out by projecting it onto its orthogonal complement. However, this orthogonal projection may harm the desired signal. Yet, when N_{rx} is large, the subspace spanned by the loop interference is nearly orthogonal to the desired signal's subspace and, hence, the orthogonal projection scheme will perform very well. The next question is: "How to project the loop interference component?" It is interesting to observe that, when N_{rx} grows large, the channel vectors of the desired signal and the loop interference become nearly orthogonal. Therefore, the ZF or the MRC receiver can act as an orthogonal projection of the loop interference. As a result, the loop interference can be reduced significantly by using large N_{rx} together with the ZF or MRC receiver. This observation is summarized in the following proposition.

Proposition 17 *Assume that the number of source-destination pairs, K, is fixed. For any finite N_{tx} or for any N_{tx}, such that $N_{\mathrm{rx}}/N_{\mathrm{tx}}$ is fixed, as $N_{\mathrm{rx}} \to \infty$, the received signal at the relay station for decoding the signal transmitted from $\mathtt{S}_k$ is given by*

$$r_k[i] \overset{a.s.}{\to} \sqrt{p_{\mathrm{S}}} x_k[i], \text{ for ZF,} \tag{19}$$

$$\frac{r_k[i]}{N_{\mathrm{rx}} \sigma^2_{\mathtt{SR},k}} \overset{a.s.}{\to} \sqrt{p_{\mathrm{S}}} x_k[i], \text{ for MRC/MRT.} \tag{20}$$

Proof: *See Appendix A.* □

The aforementioned results imply that, when N_{rx} grows to infinity, the loop interference can be canceled out. Furthermore, the interpair interference and noise effects also disappear. The received signal at the relay station after using ZF or MRC receivers includes only the desired signal and, hence, the capacity of the communication link $\mathtt{S}_k \to \mathtt{R}$ grows without bound. As a result, the system performance is limited only by the performance of the communication link $\mathtt{R} \to \mathtt{D}_k$ which does not depend on the loop interference.

3.2 Using a Large Transmit Antenna Array and Low Transmit Power ($p_{\mathrm{R}} = E_{\mathrm{R}}/N_{\mathrm{tx}}$, where E_{R} is Fixed, and $N_{\mathrm{tx}} \to \infty$)

The loop interference depends strongly on the transmit power at the relay station, p_{R}, and, hence, another way to reduce it is to use low transmit power. Unfortunately, this will also reduce the quality of the transmission link $\mathtt{R} \to \mathtt{D}_k$ and, hence, the e2e system performance will be degraded. However, with a large transmit antenna array at the relay station, we can reduce the relay transmit power while maintaining a desired quality-of-service (QoS) of the transmission link $\mathtt{R} \to \mathtt{D}_k$. This is due to the fact that, when the number of transmit antennas, N_{tx}, is large, the relay station can focus its emitted energy into the physical directions wherein the destinations are located. At the same time, the relay station can purposely avoid transmitting into physical directions where the receive antennas are located and, hence, the loop interference can be significantly reduced. Therefore, we propose to use a very large N_{tx} together with low transmit power at the relay station. With this method, the loop interference in the transmission link $\mathtt{S}_k \to \mathtt{R}$ becomes negligible, while the quality of the transmission link $\mathtt{R} \to \mathtt{D}_k$ is still fairly good. As a result, we can obtain a good e2e performance.

Proposition 18 *Assume that K is fixed and the transmit power at the relay station is $p_{\mathrm{R}} = E_{\mathrm{R}}/N_{\mathrm{tx}}$, where E_{R} is fixed regardless of N_{tx}. For any finite N_{rx}, as $N_{\mathrm{tx}} \to \infty$, the received signals at the relay station and $\mathtt{D}_k$ converge to*

$$r_k[i] \overset{a.s.}{\to} \sqrt{p_{\mathrm{S}}}\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} x_k[i] + \sqrt{p_{\mathrm{S}}} \sum_{j \neq k}^{K} \boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},j} x_j[i] + \boldsymbol{w}_k^T \boldsymbol{n}_{\mathtt{R}}[i], \text{ for both ZF and MRC/MRT,} \tag{21}$$

$$y_{\mathtt{D},k}[i] \overset{a.s.}{\to} \begin{cases} \sqrt{\frac{E_{\mathrm{R}}}{\sum_{j=1}^{K} \sigma_{\mathtt{RD},j}^{-2}}} x_k[i-d] + n_{\mathtt{D},k}[i], \text{ for ZF,} \\ \sqrt{\frac{\sigma_{\mathtt{RD},k}^4 E_{\mathrm{R}}}{\sum_{j=1}^{K} \sigma_{\mathtt{RD},j}^2}} x_k[i-d] + n_{\mathtt{D},k}[i], \text{ for MRC/MRT,} \end{cases} \tag{22}$$

respectively.

Proof: *With ZF processing, the loop interference is given by*

$$\sqrt{p_{\mathrm{R}}}\boldsymbol{W}^T \boldsymbol{G}_{\mathtt{RR}} \boldsymbol{s}[i] = \sqrt{\frac{(N_{\mathrm{tx}} - K) E_{\mathrm{R}}}{N_{\mathrm{tx}} \sum_{k=1}^{K} \sigma_{\mathtt{RD},k}^{-2}}} \boldsymbol{W}_{\mathtt{ZF}}^T \frac{\boldsymbol{G}_{\mathtt{RR}} \hat{\boldsymbol{G}}_{\mathtt{RD}}^*}{N_{\mathrm{tx}}} \left(\frac{\hat{\boldsymbol{G}}_{\mathtt{RD}}^T \hat{\boldsymbol{G}}_{\mathtt{RD}}^*}{N_{\mathrm{tx}}} \right)^{-1} \boldsymbol{x}[i-d] \overset{a.s.}{\to} 0, \text{ as } N_{\mathrm{tx}} \to \infty, \tag{23}$$

where the convergence follows the law of large numbers. Thus, we obtain (21)*. By using a similar method as in Appendix A, we can obtain* (22)*. The results for MRC/MRT processing follow a similar line of reasoning.* □

We can see that, by using a very low transmit power, i.e., scaled proportionally to $1/N_{\mathrm{tx}}$, the loop interference effect at the receive antennas is negligible [see (21)]. Although the transmit power is low, the power level of the desired signal received at each $\mathtt{D}_k$ is good enough thanks to the improved array gain, when N_{tx} grows large. At the same time, interpair interference at each $\mathtt{D}_k$ disappears due to the orthogonality between the channel vectors [see (22)]. As a result, the quality of the second hop $\mathtt{R} \to \mathtt{D}_k$ is still good enough to provide a robust overall e2e performance.

4 Achievable Rate Analysis

In this section, we derive the e2e achievable rate of the transmission link $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$ for ZF and MRC/MRT processing. The achievable rate is limited by the weakest/bottleneck link, i.e., it is equal to the minimum of the achievable rates of the transmissions from $\mathtt{S}_k$ to $\mathtt{R}$ and from $\mathtt{R}$ to $\mathtt{D}_k$ [9]. To obtain this achievable rate, we use a technique from [23]. With this technique, the received signal is rewritten as a known mean gain times the desired symbol, plus an uncorrelated effective noise whose entropy is upper-bounded by the entropy of Gaussian noise. This technique is widely used in the analysis of massive MIMO systems since: i) it yields a simplified insightful rate expression, which is basically a lower bound of what can be achieved in practice; and ii) it does not require instantaneous CSI at the destination [22, 24, 25]. The e2e achievable rate of the transmission link $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$ is given by

$$R_k = \min\left\{R_{\mathtt{SR},k}, R_{\mathtt{RD},k}\right\}, \tag{24}$$

where $R_{\mathtt{SR},k}$ and $R_{\mathtt{RD},k}$ are the achievable rates of the transmission links $\mathtt{S}_k \to \mathtt{R}$ and $\mathtt{R} \to \mathtt{D}_k$, respectively. We next compute $R_{\mathtt{SR},k}$ and $R_{\mathtt{RD},k}$. To compute $R_{\mathtt{SR},k}$, we consider (10). From (10), the received signal used for detecting $x_k[i]$ at the relay station can be written as

$$r_k[i] = \underbrace{\sqrt{p_{\mathrm{S}}}\,\mathbb{E}\left\{\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right\} x_k[i]}_{\text{desired signal}} + \underbrace{\tilde{n}_{\mathtt{R},k}[i]}_{\text{effective noise}}, \tag{25}$$

where $\tilde{n}_{\mathtt{R},k}[i]$ is considered as the effective noise, given by

$$\begin{aligned}\tilde{n}_{\mathtt{R},k}[i] \triangleq{}& \sqrt{p_{\mathrm{S}}}\Big(\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} - \mathbb{E}\left\{\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right\}\Big) x_k[i] \\ &+ \sqrt{p_{\mathrm{S}}} \sum_{j \neq k}^{K} \boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},j} x_j[i] + \sqrt{p_{\mathrm{R}}} \boldsymbol{w}_k^T \boldsymbol{G}_{\mathtt{RR}} \boldsymbol{s}[i] + \boldsymbol{w}_k^T \boldsymbol{n}_{\mathtt{R}}[i].\end{aligned} \tag{26}$$

We can see that the "desired signal" and the "effective noise" in (25) are uncorrelated. Therefore, by using the fact that the worst-case uncorrelated additive noise

is independent Gaussian noise of the same variance, we can obtain an achievable rate as

$$R_{\mathtt{SR},k} = \log_2\left(1 + \frac{p_{\mathrm{S}}\left|\mathbb{E}\left\{\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}\right\}\right|^2}{p_{\mathrm{S}}\mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}\right) + \mathtt{MP}_k + \mathtt{LI}_k + \mathtt{AN}_k}\right), \tag{27}$$

where $\mathtt{MP}_k$, $\mathtt{LI}_k$, and $\mathtt{AN}_k$ represent the multipair interference, LI, and additive noise effects, respectively, given by

$$\mathtt{MP}_k \triangleq p_{\mathrm{S}}\sum_{j\neq k}^{K}\mathbb{E}\left\{\left|\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},j}\right|^2\right\}, \tag{28}$$

$$\mathtt{LI}_k \triangleq p_{\mathrm{R}}\,\mathbb{E}\left\{\left\|\boldsymbol{w}_k^T\boldsymbol{G}_{\mathtt{RR}}\boldsymbol{A}\right\|^2\right\}, \tag{29}$$

$$\mathtt{AN}_k \triangleq \mathbb{E}\left\{\left\|\boldsymbol{w}_k\right\|^2\right\}. \tag{30}$$

To compute $R_{\mathtt{RD},k}$, we consider (12). Following a similar method as in the derivation of $R_{\mathtt{SR},k}$, we obtain

$$R_{\mathtt{RD},k} = \log_2\left(1 + \frac{p_{\mathrm{R}}\left|\mathbb{E}\left\{\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_k\right\}\right|^2}{p_{\mathrm{R}}\mathbb{V}\mathrm{ar}\left(\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_k\right) + p_{\mathrm{R}}\sum_{j\neq k}^{K}\mathbb{E}\left\{\left|\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_j\right|^2\right\} + 1}\right). \tag{31}$$

Remark 13 *The achievable rates in* (27) *and* (31) *are obtained by approximating the effective noise via an additive Gaussian noise. Since the effective noise is a sum of many terms, the central limit theorem guarantees that this is a good approximation, especially in massive MIMO systems. Hence the rate bounds in* (27) *and* (31) *are expected to be quite tight in practice.*

Remark 14 *The achievable rate* (31) *is obtained by assuming that the destination,* $\mathtt{D}_k$ *uses only statistical knowledge of the channel gains* $\left(i.e.,\ \mathbb{E}\left\{\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_k\right\}\right)$ *to decode the transmitted signals and, hence, no time, frequency, and power resources need to be allocated to the transmission of pilots for CSI acquisition. However, an interesting question is: Are our achievable rate expressions accurate predictors of the system performance? To answer this question, we compare our achievable rate* (31) *with the ergodic achievable rate of the genie receiver, i.e., the relay station knows* $\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},j}$ *and* $\boldsymbol{G}_{\mathtt{RR}}$*, and the destination* $\mathtt{D}_k$ *knows perfectly* $\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_j$*,* $j = 1, ..., K$*. For this case, the ergodic e2e achievable rate of the transmission link* $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$ *is*

$$\tilde{R}_k = \min\left\{\tilde{R}_{\mathtt{SR},k}, \tilde{R}_{\mathtt{RD},k}\right\}, \tag{32}$$

where $\tilde{R}_{\mathtt{SR},k}$ *and* $\tilde{R}_{\mathtt{RD},k}$ *are given by*

$$\tilde{R}_{\mathtt{SR},k} = \mathbb{E}\left\{\log_2\left(1+\frac{p_{\mathrm{S}}\left|\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}\right|^2}{p_{\mathrm{S}}\sum\limits_{j\neq k}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},j}\right|^2+p_{\mathrm{R}}\left\|\boldsymbol{w}_k^T\boldsymbol{G}_{\mathtt{RR}}\boldsymbol{A}\right\|^2+\|\boldsymbol{w}_k\|^2}\right)\right\}, \tag{33}$$

$$\tilde{R}_{\mathtt{RD},k} = \mathbb{E}\left\{\log_2\left(1+\frac{p_{\mathrm{R}}\left|\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_k\right|^2}{p_{\mathrm{R}}\sum_{j\neq k}^{K}\left|\boldsymbol{g}_{\mathtt{RD},k}^T\boldsymbol{a}_j\right|^2+1}\right)\right\}. \tag{34}$$

In Section 6, it is demonstrated via simulations that the performance gap between the achievable rates given by (24) *and* (32) *is rather small, especially for large* N_{rx} *and* N_{tx}*. Note that the above ergodic achievable rate in* (32) *is obtained under the assumption of perfect CSI which is idealistic in practice.*

We next provide a new approximate closed-form expression for the e2e achievable rate given by (24) for ZF, and a new exact one for MRC/MRT processing:

Theorem 3 *With ZF processing, the e2e achievable rate of the transmission link* $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$*, for a finite number of receive antennas at the relay station and* $N_{\mathrm{tx}} \gg 1$*, can be approximated as*

$$R_k \approx R_k^{\mathtt{ZF}} \triangleq \log_2\left(1+\min\left(\frac{p_{\mathrm{S}}\left(N_{\mathrm{rx}}-K\right)\sigma_{\mathtt{SR},k}^2}{p_{\mathrm{S}}\sum\limits_{j=1}^{K}\left(\beta_{\mathtt{SR},j}-\sigma_{\mathtt{SR},j}^2\right)+p_{\mathrm{R}}\sigma_{\mathtt{LI}}^2\left(1-K/N_{\mathrm{tx}}\right)+1}, \frac{N_{\mathrm{tx}}-K}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^{-2}}\frac{p_{\mathrm{R}}}{p_{\mathrm{R}}\left(\beta_{\mathtt{RD},k}-\sigma_{\mathtt{RD},k}^2\right)+1}\right)\right). \tag{35}$$

Proof: *See Appendix B.* □

Note that, the above approximation is due to the approximation of the loop interference. More specifically, to compute the loop interference term, $\mathtt{LI}_k$, we approximate $\hat{\boldsymbol{G}}_{\mathtt{RD}}^T\hat{\boldsymbol{G}}_{\mathtt{RD}}^*$ as $N_{\mathrm{tx}}\hat{\boldsymbol{D}}_{\mathtt{RD}}$. This approximation follows the law of large numbers, and, hence, becomes exact in the large-antenna limit. In fact, in Section 6, we will show that this approximation is rather tight even for a finite number of antennas.

Theorem 4 *With MRC/MRT processing, the e2e achievable rate of the transmission link* $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$*, for a finite number of antennas at the relay station, is given by*

$$R_k = R_k^{\mathtt{MR}} \triangleq \log_2\left(1 + \min\left(\frac{p_{\mathrm{S}} N_{\mathrm{rx}} \sigma_{\mathtt{SR},k}^2}{p_{\mathrm{S}} \sum_{j=1}^{K} \beta_{\mathtt{SR},j} + p_{\mathrm{R}} \sigma_{\mathtt{LI}}^2 + 1}, \frac{\sigma_{\mathtt{RD},k}^4}{\sum_{j=1}^{K} \sigma_{\mathtt{RD},j}^2} \frac{p_{\mathrm{R}} N_{\mathrm{tx}}}{p_{\mathrm{R}} \beta_{\mathtt{RD},k} + 1}\right)\right). \tag{36}$$

Proof: *See Appendix C.* □

5 Performance Evaluation

To evaluate the system performance, we consider the sum spectral efficiency. The sum spectral efficiency is defined as the sum-rate (in bits) per channel use. Let T be the length of the coherence interval (in symbols). During each coherence interval, we spend τ symbols for training, and the remaining interval is used for the payload data transmission. Therefore, the sum spectral efficiency is given by

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{A}} \triangleq \frac{T-\tau}{T} \sum_{k=1}^{K} R_k^{\mathtt{A}}, \tag{37}$$

where $\mathtt{A} \in \{\mathtt{ZF}, \mathtt{MR}\}$ corresponds to ZF and MRC/MRT processing. Note that in the case of ZF processing, $R_k^{\mathtt{ZF}}$ is an approximate result. However, in the numerical results (see Section 6.1), we show that this approximation is very tight and fairly accurate. For this reason, and without significant lack of clarity, we hereafter consider the rate results of ZF processing as exact.

From *Theorems* 3, 4, and (37), the sum spectral efficiencies of ZF and MRC/MRT processing for the full-duplex mode are, respectively, given by

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{ZF}} = \frac{T-\tau}{T} \sum_{k=1}^{K} \log_2\Bigg(1 + \min\Bigg(\frac{p_{\mathrm{S}}\left(N_{\mathrm{rx}} - K\right)\sigma_{\mathtt{SR},k}^2}{p_{\mathrm{S}} \sum_{j=1}^{K}\left(\beta_{\mathtt{SR},j} - \sigma_{\mathtt{SR},j}^2\right) + p_{\mathrm{R}}\sigma_{\mathtt{LI}}^2\left(1 - K/N_{\mathrm{tx}}\right) + 1}, \frac{N_{\mathrm{tx}} - K}{\sum_{j=1}^{K} \sigma_{\mathtt{RD},j}^{-2}} \frac{p_{\mathrm{R}}}{p_{\mathrm{R}}\left(\beta_{\mathtt{RD},k} - \sigma_{\mathtt{RD},k}^2\right) + 1}\Bigg)\Bigg), \tag{38}$$

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{MR}} = \frac{T-\tau}{T} \sum_{k=1}^{K} \log_2\left(1 + \min\left(\frac{p_{\mathrm{S}} N_{\mathrm{rx}} \sigma_{\mathtt{SR},k}^2}{p_{\mathrm{S}} \sum\limits_{j=1}^{K} \beta_{\mathtt{SR},j} + p_{\mathrm{R}}\sigma_{\mathtt{LI}}^2 + 1}, \frac{\sigma_{\mathtt{RD},k}^4}{\sum\limits_{j=1}^{K} \sigma_{\mathtt{RD},j}^2} \frac{p_{\mathrm{R}} N_{\mathrm{tx}}}{p_{\mathrm{R}} \beta_{\mathtt{RD},k} + 1}\right)\right). \tag{39}$$

5.1 Power Efficiency

In this part, we study the potential for power savings by using very large antenna arrays at the relay station.

1. *Case I*: We consider the case where p_{p} is fixed, $p_{\mathrm{S}} = E_{\mathtt{S}}/N_{\mathrm{rx}}$, and $p_{\mathrm{R}} = E_{\mathrm{R}}/N_{\mathrm{tx}}$, where $E_{\mathtt{S}}$ and E_{R} are fixed regardless of N_{rx} and N_{tx}. This case corresponds to the case where the channel estimation accuracy is fixed, and we want to investigate the potential for power saving in the data transmission phase. When N_{tx} and N_{rx} go to infinity with the same speed, the sum spectral efficiencies of ZF and MRC/MRT processing can be expressed as

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{ZF}} \to \frac{T-\tau}{T}\sum_{k=1}^{K}\log_2\left(1+\min\left(E_{\mathtt{S}}\sigma_{\mathtt{SR},k}^2, \frac{E_{\mathrm{R}}}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^{-2}}\right)\right), \tag{40}$$

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{MR}} \to \frac{T-\tau}{T}\sum_{k=1}^{K}\log_2\left(1+\min\left(E_{\mathtt{S}}\sigma_{\mathtt{SR},k}^2, \frac{\sigma_{\mathtt{RD},k}^4 E_{\mathrm{R}}}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2}\right)\right). \tag{41}$$

   The expressions in (40) and (41) show that, with large antenna arrays, we can reduce the transmitted power of each source and of the relay station proportionally to $1/N_{\mathrm{rx}}$ and $1/N_{\mathrm{tx}}$, respectively, while maintaining a given QoS. If we now assume that large-scale fading is neglected (i.e., $\beta_{\mathtt{SR},k} = \beta_{\mathtt{RD},k} = 1, \forall k$), then from (40) and (41), the asymptotic performances of ZF and MRC/MRT processing are the same and given by:

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{A}} \to \frac{T-\tau}{T} K \log_2\left(1+\sigma_1^2 \min\left(E_{\mathtt{S}}, \frac{E_{\mathrm{R}}}{K}\right)\right), \tag{42}$$

   where $\sigma_1^2 \triangleq \frac{\tau p_{\mathrm{p}}}{\tau p_{\mathrm{p}}+1}$. The sum spectral efficiency in (42) is equal to the one of K parallel single-input single-output channels with transmit power $\sigma_1^2 \min\left(E_{\mathtt{S}}, \frac{E_{\mathrm{R}}}{K}\right)$, without interference and fast fading. We see that, by using large antenna arrays, not only the transmit powers are reduced significantly, but also the sum spectral efficiency is increased K times (since all K different communication pairs are served simultaneously).

2. *Case II*: If $p_{\mathrm{p}} = p_{\mathrm{S}} = E_{\mathtt{S}}/\sqrt{N_{\mathrm{rx}}}$ and $p_{\mathrm{R}} = E_{\mathrm{R}}/\sqrt{N_{\mathrm{tx}}}$, where $E_{\mathtt{S}}$ and E_{R} are fixed regardless of N_{rx} and N_{tx}. When N_{rx} goes to infinity and $N_{\mathrm{tx}} = \kappa N_{\mathrm{rx}}$, with $\kappa > 0$, the sum spectral efficiencies converge to

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{ZF}} \to \frac{T-\tau}{T}\sum_{k=1}^{K}\log_2\left(1+\min\left(\tau E_{\mathtt{S}}^2\beta_{\mathtt{SR},k}^2, \frac{\sqrt{\kappa}\tau E_{\mathtt{S}}E_{\mathrm{R}}}{\sum_{j=1}^{K}\beta_{\mathtt{RD},j}^{-2}}\right)\right), \tag{43}$$

$$\mathcal{S}_{\mathrm{FD}}^{\mathtt{MR}} \to \frac{T-\tau}{T}\sum_{k=1}^{K}\log_2\left(1+\min\left(\tau E_{\mathtt{S}}^2\beta_{\mathtt{SR},k}^2, \frac{\sqrt{\kappa}\tau E_{\mathtt{S}}E_{\mathrm{R}}\beta_{\mathtt{RD},k}^4}{\sum_{j=1}^{K}\beta_{\mathtt{RD},j}^{-2}}\right)\right). \tag{44}$$

We see that, if the transmit powers of the uplink training and data transmission are the same, (i.e., $p_{\mathrm{p}} = p_{\mathrm{S}}$), we cannot reduce the transmit powers of each source and of the relay station as aggressively as in *Case I* where the pilot power is kept fixed. Instead, we can scale down the transmit powers of each source and of the relay station proportionally to only $1/\sqrt{N_{\mathrm{rx}}}$ and $1/\sqrt{N_{\mathrm{tx}}}$, respectively. This observation can be interpreted as, when we cut the transmitted power of each source, both the data signal and the pilot signal suffer from power reduction, which leads to the so-called "squaring effect" on the spectral efficiency [23].

5.2 Comparison between Half-Duplex and Full-Duplex Modes

In this section, we compare the performance of the half-duplex and full-duplex modes. For the half-duplex mode, two orthogonal time slots are allocated for two transmissions: sources to the relay station and the relay station to destinations [4]. The half-duplex mode does not induce loop interference at the cost of imposing a pre-log factor 1/2 on the spectral efficiency. The sum spectral efficiency of the half-duplex mode can be obtained directly from (38) and (39) by neglecting the loop interference effect. Note that, with the half-duplex mode, the sources and the relay station transmit only half of the time compared to the full-duplex mode. For fair comparison, the total energies spent in a coherence interval for both modes are set to be the same. As a result, the transmit powers of each source and of the relay station used in the half-duplex mode are double the powers used in the full-duplex mode and, hence, the sum spectral efficiencies of the half-duplex mode for ZF and

MRC/MRT processing are respectively given by[2]

$$\mathcal{S}_{\mathrm{HD}}^{\mathtt{ZF}} = \frac{T-\tau}{2T}\sum_{k=1}^{K}\log_2\left(1+\min\left(\frac{2p_{\mathrm{S}}\left(N_{\mathrm{rx}}-K\right)\sigma_{\mathtt{SR},k}^2}{2p_{\mathrm{S}}\sum_{j=1}^{K}\left(\beta_{\mathtt{SR},j}-\sigma_{\mathtt{SR},j}^2\right)+1}, \right.\right. \\ \left.\left. \frac{N_{\mathrm{tx}}-K}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^{-2}}\frac{2p_{\mathrm{R}}}{2p_{\mathrm{R}}\left(\beta_{\mathtt{RD},k}-\sigma_{\mathtt{RD},k}^2\right)+1}\right)\right), \quad (45)$$

$$\mathcal{S}_{\mathrm{HD}}^{\mathtt{MR}} = \frac{T-\tau}{2T}\sum_{k=1}^{K}\log_2\left(1+\min\left(\frac{2p_{\mathrm{S}}N_{\mathrm{rx}}\sigma_{\mathtt{SR},k}^2}{2p_{\mathrm{S}}\sum_{j=1}^{K}\beta_{\mathtt{SR},j}+1}, \frac{\sigma_{\mathtt{RD},k}^4}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2}\frac{2p_{\mathrm{R}}N_{\mathrm{tx}}}{2p_{\mathrm{R}}\beta_{\mathtt{RD},k}+1}\right)\right). \quad (46)$$

Depending on the transmit powers, channel gains, channel estimation accuracy, and the loop interference level, the full-duplex mode is preferred over the half-duplex modes and vice versa. The critical factor is the loop interference level. If all other factors are fixed, the full-duplex mode outperforms the half-duplex mode if $\sigma_{\mathtt{LI}}^2 \leq \sigma_{\mathtt{LI},0}^2$, where $\sigma_{\mathtt{LI},0}^2$ is the root of $\mathcal{S}_{\mathrm{FD}}^{\mathtt{ZF}} = \mathcal{S}_{\mathrm{HD}}^{\mathtt{ZF}}$ for the ZF processing or the root of $\mathcal{S}_{\mathrm{FD}}^{\mathtt{MR}} = \mathcal{S}_{\mathrm{HD}}^{\mathtt{MR}}$ for the MRC/MRT processing.

From the above observation, we propose to use a hybrid relaying mode as follows:

$$\mathtt{Hybrid\ Relaying\ Mode} = \begin{cases} \mathtt{Full-Duplex}, \ \mathtt{if}\ \mathcal{S}_{\mathrm{FD}}^{\mathtt{A}} \geq \mathcal{S}_{\mathrm{HD}}^{\mathtt{A}} \\ \mathtt{Half-Duplex}, \ \mathtt{otherwise}. \end{cases}$$

Note that, with hybrid relaying, the relaying mode is chosen for each large-scale fading realization.

5.3 Power Allocation

In previous sections, we assumed that the transmit powers of all users are the same. The system performance can be improved by optimally allocating different powers to different sources. Thus, in this section, we assume that the transmit powers of different sources are different. We assume that the design for training phase is done in advance, i.e., the training duration, τ, and the pilot power, p_{p}, were determined. We are interested in designing a power allocation algorithm in

[2]Here, we assume that the relay station in the half-duplex mode employs the same number of transmit and receive antennas as in the full-duplex mode. This assumption corresponds to the "RF chains conserved" condition, where an equal number of total RF chains are assumed [11, Section III]. Note that, in order to receive the transmitted signals from the destinations during the channel estimation phase, additional "receive RF chains" have to be used in the transmit array for both the full-duplex and half-duplex cases. The comparison between half-duplex and full-duplex modes can be also performed with the "number of antennas preserved" condition, where the number of antennas at the relay station used in the half-duplex mode is equal to the total number of transmit and receive antennas used in the FD mode, i.e., is equal to $N_{\mathrm{tx}} + N_{\mathrm{rx}}$. However, the cost of the required RF chains is significant as opposed to adding an extra antenna. Thus, we choose the "RF chains conserved" condition for our comparison.

the data transmission phase that maximizes the energy efficiency, for each large-scale realization, subject to a given sum spectral efficiency and the constraints of maximum powers transmitted from sources and the relay station. The energy efficiency (in bits/Joule) is defined as the sum spectral efficiency divided by the total transmit power. Let the transmit power of the kth source be $p_{\mathtt{S},k}$. Therefore, the energy efficiency of the full-duplex mode is given by

$$\mathtt{EE}^{\mathtt{A}} \triangleq \frac{\mathcal{S}_{\mathrm{FD}}^{\mathtt{A}}}{\frac{T-\tau}{T}\left(\sum_{k=1}^{K} p_{\mathtt{S},k} + p_{\mathrm{R}}\right)}. \tag{47}$$

Mathematically, the optimization problem can be formulated as

$$\begin{array}{ll} \text{maximize} & \mathtt{EE}^{\mathtt{A}} \\ \text{subject to} & \mathcal{S}_{\mathrm{FD}}^{\mathtt{A}} = \mathcal{S}_{0}^{\mathtt{A}} \\ & 0 \leq p_{\mathtt{S},k} \leq p_0, k = 1, ..., K \\ & 0 \leq p_{\mathrm{R}} \leq p_1 \end{array} \tag{48}$$

where $\mathcal{S}_{0}^{\mathtt{A}}$ is a required sum spectral efficiency, while p_0 and p_1 are the peak power constraints of $p_{\mathtt{S},k}$ and p_{R}, respectively.

From (38), (39), and (47), the optimal power allocation problem in (48) can be rewritten as

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{K} p_{\mathtt{S},k} + p_{\mathrm{R}} \\ \text{subject to} & \\ & \frac{T-\tau}{T}\sum\limits_{k=1}^{K} \log_2\left(1+\min\left\{\frac{a_k p_{\mathtt{S},k}}{\sum\limits_{j=1}^{K} b_j p_{\mathtt{S},j}+c_k p_{\mathrm{R}}+1}, \frac{d_k p_{\mathrm{R}}}{e_k p_{\mathrm{R}}+1}\right\}\right) = \mathcal{S}_{0}^{\mathtt{A}} \\ & 0 \leq p_{\mathtt{S},k} \leq p_0, k = 1, ..., K \\ & 0 \leq p_{\mathrm{R}} \leq p_1 \end{array} \tag{49}$$

where a_k, b_k, c_k, d_k, and e_k are constant values (independent of the transmit powers) which are different for ZF and MRC/MRT processing. More precisely,

- For ZF: $a_k = (N_{\mathrm{rx}} - K)\,\sigma_{\mathtt{SR},k}^2$, $b_k = \beta_{\mathtt{SR},k} - \sigma_{\mathtt{SR},k}^2$, $c_k = \sigma_{\mathtt{LI}}^2\,(1 - K/N_{\mathrm{tx}})$, $d_k = \frac{N_{\mathrm{tx}}-K}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^{-2}}$, and $e_k = \beta_{\mathtt{RD},k} - \sigma_{\mathtt{RD},k}^2$.
- For MRC/MRT: $a_k = N_{\mathrm{rx}}\sigma_{\mathtt{SR},k}^2$, $b_k = \beta_{\mathtt{SR},k}$, $c_k = \sigma_{\mathtt{LI}}^2$, $d_k = \frac{\sigma_{\mathtt{RD},k}^4}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2} N_{\mathrm{tx}}$, and $e_k = \beta_{\mathtt{RD},k}$.

The problem (49) is equivalent to

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{K} p_{\mathtt{S},k} + p_{\mathrm{R}} \\ \text{subject to} & \frac{T-\tau}{T}\sum_{k=1}^{K}\log_2(1+\gamma_k) = \mathcal{S}_{0}^{\mathtt{A}} \\ & \gamma_k \leq \frac{a_k p_{\mathtt{S},k}}{\sum_{j=1}^{K} b_j p_{\mathtt{S},j}+c_k p_{\mathrm{R}}+1}, k = 1, ..., K \\ & \gamma_k \leq \frac{d_k p_{\mathrm{R}}}{e_k p_{\mathrm{R}}+1}, k = 1, ..., K \\ & 0 \leq p_{\mathtt{S},k} \leq p_0, k = 1, ..., K \\ & 0 \leq p_{\mathrm{R}} \leq p_1. \end{array} \tag{50}$$

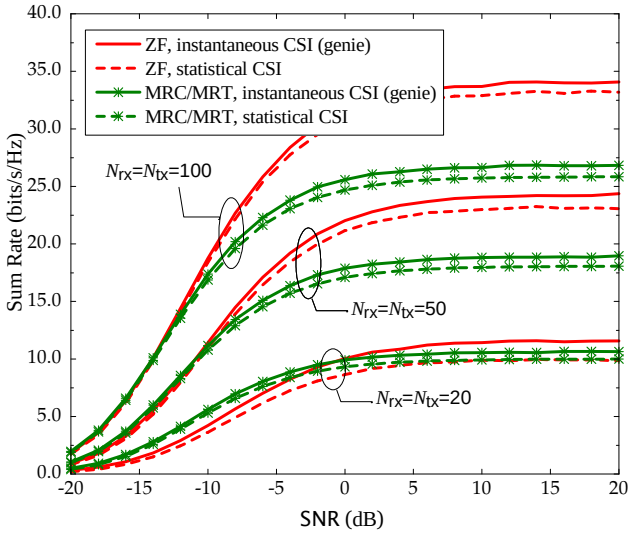


Figure 2: Sum rate versus SNR for ZF and MRC/MRT processing ($K = 10$, $\tau = 2K$, and $\sigma^2_{\text{LI}} = 1$).

Since a_k, b_k, c_k, d_k, and e_k are positive, (50) can be equivalently written as

$$\begin{aligned} &\text{minimize} \quad \textstyle\sum_{k=1}^{K} p_{\text{S},k} + p_{\text{R}} \\ &\text{subject to} \quad \textstyle\prod_{k=1}^{K} (1+\gamma_k) = 2^{\frac{T\mathcal{S}^{\text{A}}_0}{T-\tau}} \\ &\quad \textstyle\sum_{j=1}^{K} \frac{b_j}{a_k} p_{\text{S},j} \gamma_k p_{\text{S},k}^{-1} + \frac{c_k}{a_k} p_{\text{R}} \gamma_k p_{\text{S},k}^{-1} + \frac{1}{a_k} \gamma_k p_{\text{S},k}^{-1} \leq 1, \forall k \\ &\quad \tfrac{e_k}{d_k} \gamma_k + \tfrac{1}{d_k} \gamma_k p_{\text{R}}^{-1} \leq 1, k = 1, ..., K \\ &\quad 0 \leq p_{\text{S},k} \leq p_0, k = 1, ..., K, \\ &\quad 0 \leq p_{\text{R}} \leq p_1. \end{aligned} \tag{51}$$

We can see that the objective function and the inequality constraints are posynomial functions. If the equality constraint is a monomial function, the problem (51) becomes a GP which can be reformulated as a convex problem, and can be solved efficiently by using convex optimization tools, such as CVX [26]. However, the equality constraint in (51) is a posynomial function, so we cannot solve (51) directly using convex optimization tools. Yet, by using the technique in [27], we can efficiently find an approximate solution of (51) by solving a sequence of GPs. More precisely, from [27, Lemma 1], we can use $\kappa_k \gamma_k^{\eta_k}$ to approximate $1 + \gamma_k$ near a point $\hat{\gamma}_k$, where $\eta_k \triangleq \hat{\gamma}_k (1 + \hat{\gamma}_k)^{-1}$ and $\kappa_k \triangleq \hat{\gamma}_k^{-\eta_k} (1 + \hat{\gamma}_k)$. As a consequence, near a point $\hat{\gamma}_k$, the left hand side of the equality constraint can be approximated as

$$\prod_{k=1}^{K} (1 + \gamma_k) \approx \prod_{k=1}^{K} \kappa_k \gamma_k^{\eta_k}, \tag{52}$$

which is a monomial function. Thus, by using the local approximation given by (52), the optimization problem (51) can be approximated by a GP. By using a similar technique as in [27], we formulate the following algorithm to solve (51):

Algorithm 4 (Successive approximation algorithm for (51)**)**

1. Initialization*: set* $i = 1$*, choose the initial values of* γ_k *as* $\gamma_{k,1}$*,* $k = 1, ..., K$*. Define a tolerance* ϵ*, the maximum number of iterations* L*, and parameter* α*.*

2. Iteration i*: compute* $\eta_{k,i} = \gamma_{k,i}\left(1+\gamma_{k,i}\right)^{-1}$ *and* $\kappa_{k,i} = \gamma_{k,i}^{-\eta_{k,i}}\left(1+\gamma_{k,i}\right)$*. Then, solve the GP:*

$$\begin{array}{ll} \textit{minimize} & \sum_{k=1}^{K} p_{\mathrm{S},k} + p_{\mathrm{R}} \\ \textit{subject to} & \prod_{k=1}^{K} \kappa_{k,i}\gamma_k^{\eta_{k,i}} = 2^{\frac{T\mathcal{S}_0^{\mathbb{A}}}{T-\tau}} \\ & \sum_{j=1}^{K} \frac{b_j}{a_k} p_{\mathrm{S},j}\gamma_k p_{\mathrm{S},k}^{-1} + \frac{c_k}{a_k} p_{\mathrm{R}}\gamma_k p_{\mathrm{S},k}^{-1} + \frac{1}{a_k}\gamma_k p_{\mathrm{S},k}^{-1} \leq 1, \forall k \\ & \frac{e_k}{d_k}\gamma_k + \frac{1}{d_k}\gamma_k p_{\mathrm{R}}^{-1} \leq 1, k = 1, ..., K \\ & 0 \leq p_{\mathrm{S},k} \leq p_0, k = 1, ..., K,\ 0 \leq p_{\mathrm{R}} \leq p_1 \\ & \alpha^{-1}\gamma_{k,i} \leq \gamma_k \leq \alpha\gamma_{k,i} \end{array}$$

Let γ_k^**,* $k = 1, ..., K$ *be the solutions.*

3. *If* $\max_k |\gamma_{k,i} - \gamma_k^*| < \epsilon$ *or* $i = L \rightarrow$ *Stop. Otherwise, go to step 4.*

4. *Set* $i = i + 1$*,* $\gamma_{k,i} = \gamma_k^*$*, go to step 2.*

Note that the parameter $\alpha > 1$ is used to control the approximation accuracy in (52). If α is close to 1, the accuracy is high, but the convergence speed is low and vice versa if α is large. As discussed in [27], $\alpha = 1.1$ offers a good accuracy and convergence speed tradeoff.

6 Numerical Results

In all illustrative examples, we choose the length of the coherence interval to be $T = 200$ (symbols), the number of communication pairs $K = 10$, the training length $\tau = 2K$, and $N_{\mathrm{tx}} = N_{\mathrm{rx}}$. Furthermore, we define $\mathsf{SNR} \triangleq p_{\mathrm{S}}$.

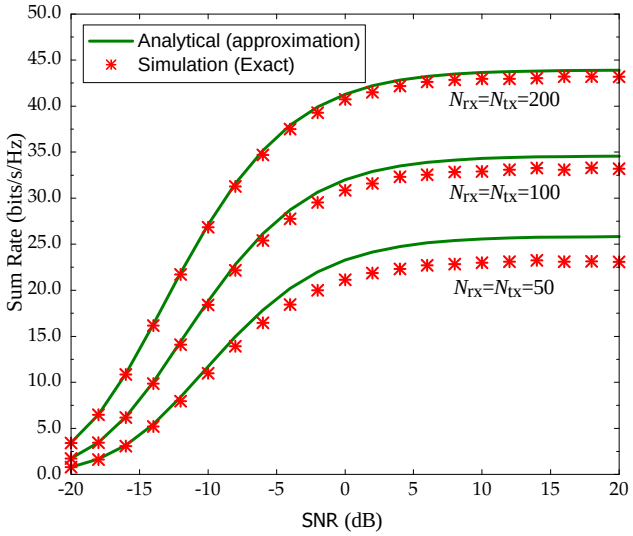


Figure 3: Sum rate versus SNR for ZF processing ($K = 10$, $\tau = 2K$, and $\sigma^2_{\text{LI}} = 1$).

6.1 Validation of Achievable Rate Results

In this subsection, we evaluate the validity of our achievable rate given by (24) as well as the approximation used to derive the closed-form expression given in Theorem 3. We choose the loop interference level $\sigma^2_{\text{LI}} = 1$. We assume that $p_{\text{P}} = p_{\text{S}}$, and that the total transmit power of the K sources is equal to the transmit power of the relay station, i.e., $p_{\text{R}} = Kp_{\text{S}}$.

We first compare our achievable rate given by (24), where the destination uses the statistical distributions of the channels (i.e., the means of channel gains) to detect the transmitted signal, with the one obtained by (32), where we assume that there is a genie receiver (instantaneous CSI) at the destination. Figure 2 shows the sum rate versus SNR for ZF and MRC/MRT processing. The dashed lines represent the sum rates obtained numerically from (24), while the solid lines represent the ergodic sum rates obtained from (32). We can see that the relative performance gap between the cases with instantaneous (genie) and statistical CSI at the destinations is small. For example, with $N_{\text{rx}} = N_{\text{tx}} = 50$, at SNR = 5dB, the sum-rate gaps are 0.65 bits/s/Hz and 0.9 bits/s/Hz for MRC/MRT and ZF processing, respectively. This implies that using the mean of the effective channel gain for signal detection is fairly reasonable, and the achievable rate given in (24) is a good predictor of the system performance.

Next, we evaluate the validity of the approximation given by (35). Figure 3 shows the sum rate versus SNR for different numbers of transmit (receive) antennas. The "Analytical (approximation)" curves are obtained by using Theorem 3, and the "Simulation (exact)" curves are generated from the outputs of a Monte-Carlo simulator

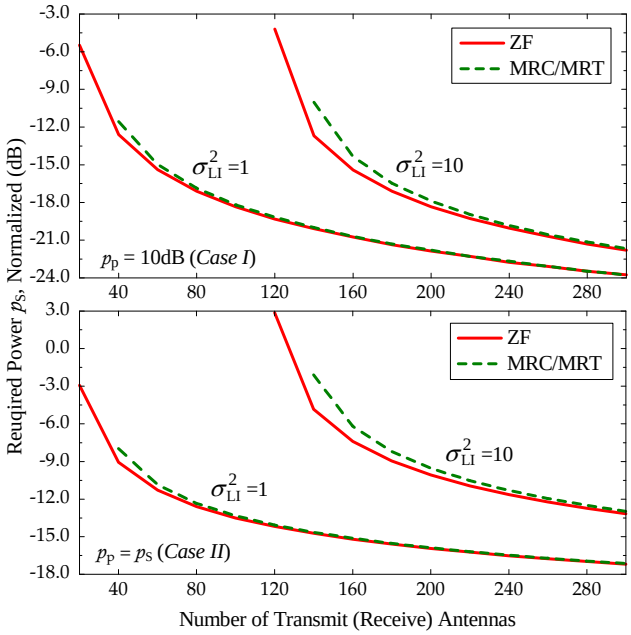


Figure 4: Transmit power, p_{S}, required to achieve 1 bit/s/Hz per user for ZF and MRC/MRT processing ($K = 10$, $\tau = 2K$, and $p_{\mathrm{R}} = Kp_{\mathrm{S}}$).

using (24), (27), and (31). We can see that the proposed approximation is very tight, especially for large antenna arrays.

6.2 Power Efficiency

We now examine the power efficiency of using large antenna arrays for two cases: p_{P} is fixed (*Case I*) and $p_{\mathrm{P}} = p_{\mathrm{S}}$ (*Case II*). We will examine how much transmit power is needed to reach a predetermined sum spectral efficiency. We set $p_{\mathrm{R}} = Kp_{\mathrm{S}}$ and $\beta_{\mathtt{SR},k} = \beta_{\mathtt{RD},k} = 1$, $k = 1, 2, ..., K$. Figure 4 shows the required transmit power, p_{S}, to achieve 1 bit/s/Hz per communication pair. We can see that when the number of antennas increases, the required transmit powers are significantly reduced. As predicted by the analysis, in the large-antenna regime, we can cut back the power by approximately 3dB and 1.5dB by doubling the number of antennas for *Case I* and *Case II*, respectively. When the loop interference is high and the number of antennas is moderate, the power efficiency can benefit more by increasing the number of antennas. For instance, for $\sigma^2_{\mathtt{LI}} = 10$, increasing the number of antennas from 120 to 240 yields a power reduction of 15dB and 13dB for *Case I* and *Case II*, respectively. Regarding the loop interference effect, when $\sigma^2_{\mathtt{LI}}$ increases, we need more transmit power. However, when $\sigma^2_{\mathtt{LI}}$ is high and the number of antennas is small, even if we use infinite transmit power, we cannot achieve a required sum

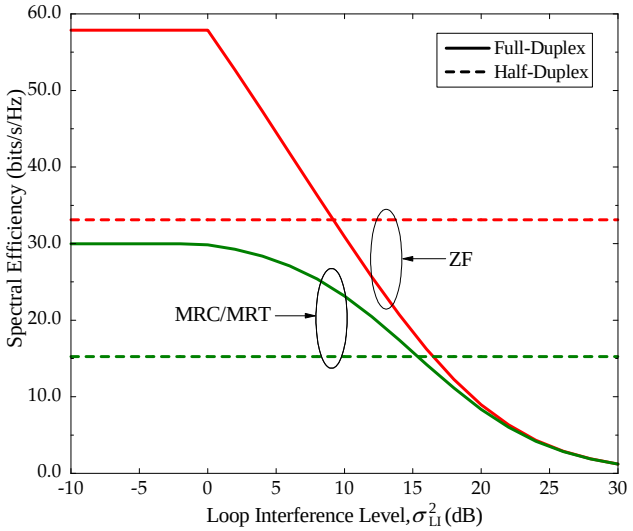


Figure 5: Sum spectral efficiency versus the loop interference levels for half-duplex and full-duplex relaying ($K = 10$, $\tau = 2K$, $p_{\mathrm{R}} = p_{\mathrm{p}} = p_{\mathrm{S}} = 10$dB, and $N_{\mathrm{tx}} = N_{\mathrm{rx}} = 100$).

spectral efficiency. Instead of this, we can add more antennas to reduce the loop interference effect and achieve the required QoS. Furthermore, when the number of antennas is large, the difference in performance between ZF and MRC/MRT processing is negligible.

6.3 Full-Duplex Vs. Half-Duplex, Hybrid Relaying Mode

Firstly, we compare the performance between half-duplex and full-duplex relaying for different loop interference levels, σ^2_{LI}. We choose $p_{\mathrm{R}} = p_{\mathrm{p}} = p_{\mathrm{S}} = 10$dB, $\beta_{\mathrm{SR},k} = \beta_{\mathrm{RD},k} = 1$, $\forall k$, and $N_{\mathrm{rx}} = N_{\mathrm{tx}} = 100$. Figure 5 shows the sum spectral efficiency versus the loop interference levels for ZF and MRC/MRT. As expected, at low σ^2_{LI}, full-duplex relaying outperforms half-duplex relaying. This gain is due to the larger pre-log factor (one) of the full-duplex mode. However, when σ^2_{LI} is high, loop interference dominates the system performance of the full-duplex mode and, hence, the performance of the half-duplex mode is superior. In this case, by using larger antenna arrays at the relay station, we can reduce the effect of the loop interference and exploit the larger pre-log factor of the full-duplex mode. This fact is illustrated in Fig. 6 where the sum spectral efficiency is represented as a function of the number of antennas, at $\sigma^2_{\mathrm{LI}} = 10$dB.

We next consider a more practical scenario that incorporates small-scale fading and large-scale fading. The large-scale fading is modeled by path loss, shadow fading,

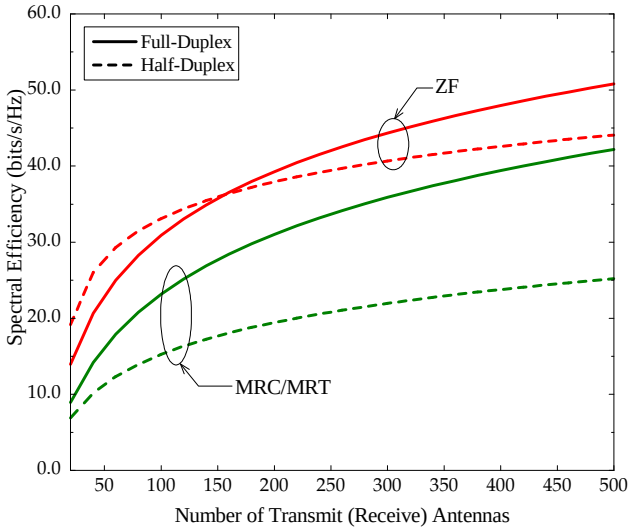


Figure 6: Sum spectral efficiency versus the number of transmit (receive) antennas for half-duplex and full-duplex relaying ($K = 10$, $\tau = 2K$, $p_{\mathrm{R}} = p_{\mathrm{p}} = p_{\mathrm{S}} = 10$dB, and $\sigma^2_{\mathtt{LI}} = 10$dB).

and random source and destination locations. More precisely, the large-scale fading $\beta_{\mathtt{SR},k}$ is

$$\beta_{\mathtt{SR},k} = \frac{z_{\mathtt{SR},k}}{1 + (\ell_k/\ell_0)^{\nu}},$$

where $z_{\mathtt{SR},k}$ represents a log-normal random variable with standard deviation of σdB, ν is the path loss exponent, ℓ_k denotes the distance between $\mathtt{S}_k$ and the receive array of the relay station, and ℓ_0 is a reference distance. We use the same channel model for $\beta_{\mathtt{RD},k}$.

We assume that all sources and destinations are located at random inside a disk with a diameter of 1000m so that ℓ_k is uniformly distributed between 0 and 500. For our simulation, we choose $\sigma = 8$dB, $\nu = 3.8$, $\ell_0 = 200$m, which are typical values in an urban cellular environment [28]. Furthermore, we choose $N_{\mathrm{rx}} = N_{\mathrm{tx}} = 200$, $p_{\mathrm{R}} = p_{\mathrm{p}} = p_{\mathrm{S}} = 10$dB, and $\sigma^2_{\mathtt{LI}} = 10$dB. Figure 7 illustrates the cumulative distributions of the sum spectral efficiencies for the half-duplex, full-duplex, and hybrid modes. The ZF processing outperforms the MRC/MRT processing in this example, and the sum spectral efficiency of MRC/MRT processing is more concentrated around its mean compared to the ZF processing. Furthermore, we can see that, for MRC/MRT, the full-duplex mode is always better than the half-duplex mode, while for ZF, depending on the large-scale fading, full-duplex can be better than half-duplex relaying and vice versa. In this example, it is also shown that hybrid relaying provides a large gain for the ZF processing case.

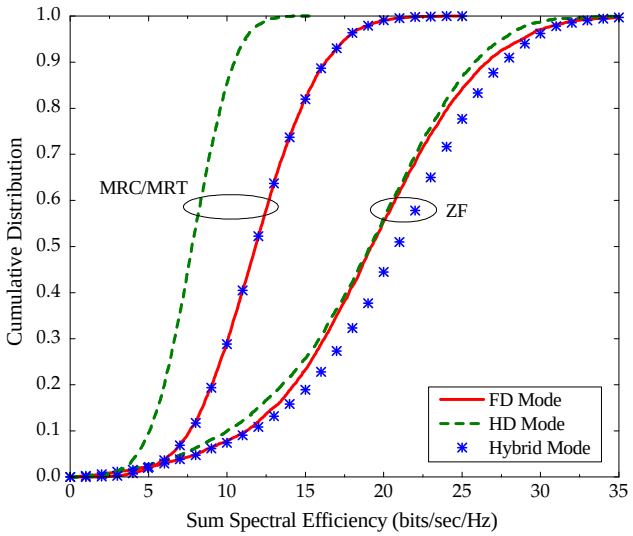


Figure 7: Cumulative distribution of the sum spectral efficiency for half-duplex, full-duplex, and hybrid relaying ($K = 10$, $\tau = 2K$, $p_{\mathrm{R}} = p_{\mathrm{p}} = p_{\mathrm{S}} = 10$dB, and $\sigma^2_{\mathrm{LI}} = 10$dB).

6.4 Power Allocation

In the following, we will examine the energy efficiency versus the sum spectral efficiency under the optimal power allocation, as outlined in Section 5.3. In this example, we choose $p_{\mathrm{p}} = 10$dB and $\sigma^2_{\mathrm{LI}} = 10$dB. Furthermore, the large-scale fading matrices are chosen as follows:

$$\boldsymbol{D}_{\mathrm{SR}} = \mathrm{diag}\left[0.749\ 0.246\ 0.125\ 0.635\ 4.468\ 0.031\ 0.064\ 0.257\ 0.195\ 0.315\right],$$
$$\boldsymbol{D}_{\mathrm{RD}} = \mathrm{diag}\left[0.070\ 0.121\ 0.134\ 0.209\ 0.198\ 0.184\ 0.065\ 0.051\ 0.236\ 1.641\right].$$

Note that, the above large-scale coefficients are obtained by taking one snapshot of the practical setup for Fig. 7.

Figure 8 shows the energy efficiency versus the sum spectral efficiency under uniform and optimal power allocation. The "uniform power allocation" curves correspond to the case where all sources and the relay station use their maximum powers, i.e., $p_{\mathrm{S},k} = p_0$, $\forall k = 1, ..., K$, and $p_{\mathrm{R}} = p_1$. The "optimal power allocation" curves are obtained by using the optimal power allocation scheme via Algorithm 4. The initial values of Algorithm 4 are chosen as follows: $\epsilon = 0.01$, $L = 5$, $\alpha = 1.1$, and $\gamma_{k,1} = \min\left\{\frac{a_k p_0}{p_0 \sum_{j=1}^{K} b_j + c_k p_1 + 1}, \frac{d_k p_1}{e_k p_1 + 1}\right\}$ which correspond to the uniform power allocation case. We can see that with optimal power allocation, the system performance improves significantly, especially at low spectral efficiencies. For example, with

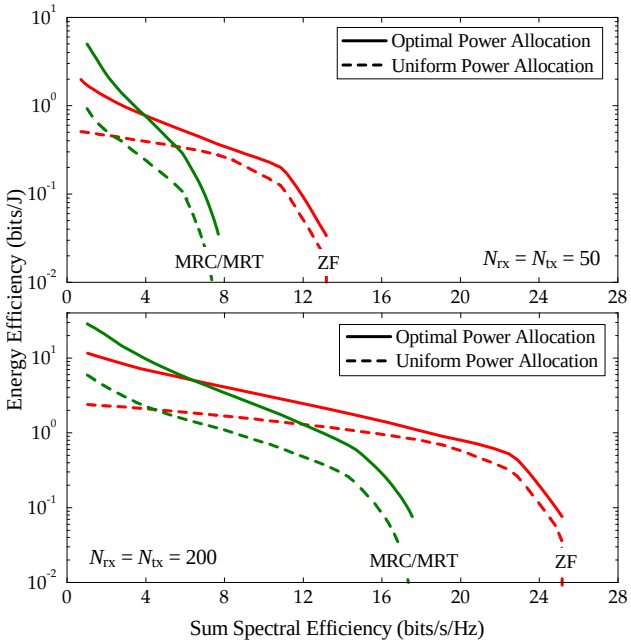


Figure 8: Energy efficiency versus sum spectral efficiency for ZF and MRC/MRT ($K = 10$, $\tau = 2K$, $p_{\mathrm{p}} = 10$dB, and $\sigma^2_{\mathrm{LI}} = 10$dB).

$N_{\mathrm{rx}} = N_{\mathrm{tx}} = 200$, to achieve the same sum spectral efficiency of 10bits/s/Hz, optimal power allocation can improve the energy efficiency by factors of 2 and 3 for ZF and MRC/MRT processing, respectively, compared to the case of no power allocation. This manifests that MRC/MRT processing benefits more from power allocation. Furthermore, at low spectral efficiencies, MRC/MRT performs better than ZF and vice versa at high spectral efficiencies. The results also demonstrate the significant benefit of using large antenna arrays at the relay station. With ZF processing, by increasing the number of antennas from 50 to 200, the energy efficiency can be increased by 14 times, when each pair has a throughput of about one bit per channel use.

7 Conclusion

In this paper, we introduced and analyzed a multipair full-duplex relaying system, where the relay station is equipped with massive arrays, while each source and destination have a single antenna. We assume that the relay station employs ZF and MRC/MRT to process the signals. Our analysis took the energy and bandwidth costs of channel estimation into account. We show that, by using massive arrays at the relay station, loop interference can be canceled out. Furthermore, the interpair

interference and noise disappear. As a result, massive MIMO can increase the sum spectral efficiency by $2K$ times compared to the conventional orthogonal half-duplex relaying, and simultaneously reduce the transmit power significantly. We derived closed-form expressions for the achievable rates and compared the performance of the full-duplex and half-duplex modes. In addition, we proposed a power allocation scheme which chooses optimally the transmit powers of the K sources and relay station to maximize the energy efficiency, subject to a given sum spectral efficiency and peak power constraints. With the proposed optimal power allocation, the energy efficiency can be significantly improved.

Appendix

A Proof of Proposition 17

1. For ZF processing:

   Here, we first provide the proof for ZF processing. From (7) and (13), we have

$$\begin{aligned}\sqrt{p_{\mathrm{S}}}\boldsymbol{W}^{T}\boldsymbol{G}_{\mathtt{SR}}\boldsymbol{x}\left[i\right] &= \sqrt{p_{\mathrm{S}}}\boldsymbol{W}_{\mathtt{ZF}}^{T}\left(\hat{\boldsymbol{G}}_{\mathtt{SR}}+\boldsymbol{\mathcal{E}}_{\mathtt{SR}}\right)\boldsymbol{x}\left[i\right] \\ &= \sqrt{p_{\mathrm{S}}}\boldsymbol{x}\left[i\right]+\sqrt{p_{\mathrm{S}}}\boldsymbol{W}_{\mathtt{ZF}}^{T}\boldsymbol{\mathcal{E}}_{\mathtt{SR}}\boldsymbol{x}\left[i\right]. \end{aligned} \tag{53}$$

   By using the law of large numbers, we obtain[3]

$$\begin{aligned}\sqrt{p_{\mathrm{S}}}\boldsymbol{W}_{\mathtt{ZF}}^{T}\boldsymbol{\mathcal{E}}_{\mathtt{SR}}\boldsymbol{x}\left[i\right] &= \sqrt{p_{\mathrm{S}}}\left(\frac{\hat{\boldsymbol{G}}_{\mathtt{SR}}^{H}\hat{\boldsymbol{G}}_{\mathtt{SR}}}{N_{\mathrm{rx}}}\right)^{-1}\frac{\hat{\boldsymbol{G}}_{\mathtt{SR}}^{H}\boldsymbol{\mathcal{E}}_{\mathtt{SR}}}{N_{\mathrm{rx}}}\boldsymbol{x}\left[i\right] \\ &\overset{a.s.}{\to} 0,\ \text{as } N_{\mathrm{rx}}\to\infty. \end{aligned} \tag{54}$$

   Therefore, as $N_{\mathrm{rx}}\to\infty$, we have

$$\sqrt{p_{\mathrm{S}}}\boldsymbol{W}^{T}\boldsymbol{G}_{\mathtt{SR}}\boldsymbol{x}\left[i\right]\overset{a.s.}{\to}\sqrt{p_{\mathrm{S}}}\boldsymbol{x}\left[i\right]. \tag{55}$$

   From (55), we can see that, when N_{rx} goes to infinity, the desired signal converges to a deterministic value, while multi-pair interference is cancelled out. More precisely, as $N_{\mathrm{rx}}\to\infty$,

$$\sqrt{p_{\mathrm{S}}}\boldsymbol{w}_{k}^{T}\boldsymbol{g}_{\mathtt{SR},k}x_{k}\left[i\right]\overset{a.s.}{\to}\sqrt{p_{\mathrm{S}}}x_{k}\left[i\right], \tag{56}$$

$$\sqrt{p_{\mathrm{S}}}\boldsymbol{w}_{k}^{T}\boldsymbol{g}_{\mathtt{SR},j}x_{j}\left[i\right]\overset{a.s.}{\to}0,\ \forall j\neq k. \tag{57}$$

[3] The law of large numbers: Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be mutually independent $n\times 1$ vectors. Suppose that the elements of $\boldsymbol{p}$ are i.i.d. zero-mean random variables with variance σ_p^2, and that the elements of $\boldsymbol{q}$ are i.i.d. zero-mean random variables with variance σ_q^2. Then, we have

$$\frac{1}{n}\boldsymbol{p}^{H}\boldsymbol{p}\overset{a.s.}{\to}\sigma_p^2,\ \text{and}\ \ \frac{1}{n}\boldsymbol{p}^{H}\boldsymbol{q}\overset{a.s.}{\to}0,\ \text{as } n\to\infty.$$

Next, we consider the loop interference term. With ZF processing, we have

$$\sqrt{p_{\mathrm{R}}}\boldsymbol{W}^T\boldsymbol{G}_{\mathtt{RR}}\boldsymbol{s}\left[i\right]=\alpha_{\mathtt{ZF}}\sqrt{p_{\mathrm{R}}}\left(\frac{\hat{\boldsymbol{G}}_{\mathtt{SR}}^H\hat{\boldsymbol{G}}_{\mathtt{SR}}}{N_{\mathrm{rx}}}\right)^{-1}\frac{\hat{\boldsymbol{G}}_{\mathtt{SR}}^H\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{G}}_{\mathtt{RD}}^*}{N_{\mathrm{rx}}N_{\mathrm{tx}}}\left(\frac{\hat{\boldsymbol{G}}_{\mathtt{RD}}^T\hat{\boldsymbol{G}}_{\mathtt{RD}}^*}{N_{\mathrm{tx}}}\right)^{-1}\boldsymbol{x}\left[i-d\right]. \quad (58)$$

If N_{tx} is fixed, then it is obvious that $\sqrt{p_{\mathrm{R}}}\boldsymbol{W}^T\boldsymbol{G}_{\mathtt{RR}}\boldsymbol{s}\left[i\right]\to 0$, as $N_{\mathrm{rx}}\to\infty$. We now consider the case where N_{tx} and N_{rx} tend to infinity with a fixed ratio. The (m,n)th element of the $K\times K$ matrix $\alpha_{\mathtt{ZF}}\frac{\hat{\boldsymbol{G}}_{\mathtt{SR}}^H\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{G}}_{\mathtt{RD}}^*}{N_{\mathrm{rx}}N_{\mathrm{tx}}}$ can be written as

$$\alpha_{\mathtt{ZF}}\frac{\hat{\boldsymbol{g}}_{\mathtt{SR},m}^H\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{g}}_{\mathtt{RD},n}^*}{N_{\mathrm{rx}}N_{\mathrm{tx}}}=\sqrt{\frac{N_{\mathrm{tx}}-K}{N_{\mathrm{tx}}\sum_{k=1}^K\sigma_{\mathtt{RD},k}^{-2}}}\frac{1}{N_{\mathrm{rx}}}\hat{\boldsymbol{g}}_{\mathtt{SR},m}^H\frac{\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{g}}_{\mathtt{RD},n}^*}{\sqrt{N_{\mathrm{tx}}}}. \quad (59)$$

We can see that the vector $\frac{\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{g}}_{\mathtt{RD},n}^*}{\sqrt{N_{\mathrm{tx}}}}$ includes i.i.d. zero-mean random variables with variance $\sigma_{\mathtt{RD},n}^2\sigma_{\mathtt{LI}}^2$. This vector is independent of the vector $\hat{\boldsymbol{g}}_{\mathtt{SR},m}$. Thus, by using the law of large numbers, we can obtain

$$\alpha_{\mathtt{ZF}}\frac{\hat{\boldsymbol{g}}_{\mathtt{SR},m}^H\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{g}}_{\mathtt{RD},n}^*}{N_{\mathrm{rx}}N_{\mathrm{tx}}}\overset{a.s.}{\to}0,\ \text{as}\ N_{\mathrm{rx}}\to\infty, N_{\mathrm{rx}}/N_{\mathrm{tx}}\ \text{is fixed}. \quad (60)$$

Therefore, the loop interference converges to 0 when N_{rx} grows without bound. Similarly, we can show that

$$\boldsymbol{W}^T\boldsymbol{n}_{\mathtt{R}}\left[i\right]\overset{a.s.}{\to}0. \quad (61)$$

Substituting (56), (57), (60), and (61) into (10), we arrive at (19).

2. For MRC/MRT processing:

We next provide the proof for MRC/MRT processing. From (7) and (16), and by using the law of large numbers, as $N_{\mathrm{rx}}\to\infty$, we have that

$$\frac{1}{N_{\mathrm{rx}}}\sqrt{p_{\mathrm{S}}}\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}x_k\left[i\right]=\frac{1}{N_{\mathrm{rx}}}\sqrt{p_{\mathrm{S}}}\hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\boldsymbol{g}_{\mathtt{SR},k}x_k\left[i\right]\overset{a.s.}{\to}\sqrt{p_{\mathrm{S}}}\sigma_{\mathtt{SR},k}^2x_k\left[i\right], \quad (62)$$

$$\frac{1}{N_{\mathrm{rx}}}\sqrt{p_{\mathrm{S}}}\boldsymbol{w}_K^T\boldsymbol{g}_{\mathtt{SR},j}x_k\left[j\right]=\frac{1}{N_{\mathrm{rx}}}\sqrt{p_{\mathrm{S}}}\hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\boldsymbol{g}_{\mathtt{SR},j}x_k\left[j\right]\overset{a.s.}{\to}0,\ \forall j\neq k. \quad (63)$$

We next consider the loop interference term. For any finite N_{tx}, or any N_{tx} where $N_{\mathrm{rx}}/N_{\mathrm{tx}}$ is fixed, as $N_{\mathrm{rx}}\to\infty$, we have

$$\frac{1}{N_{\mathrm{rx}}}\sqrt{p_{\mathrm{R}}}\boldsymbol{W}^T\boldsymbol{G}_{\mathtt{RR}}\boldsymbol{s}\left[i\right]=\alpha_{\mathtt{MRT}}\sqrt{p_{\mathrm{R}}}\frac{\hat{\boldsymbol{G}}_{\mathtt{SR}}^H\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{G}}_{\mathtt{RD}}^*}{N_{\mathrm{rx}}}\boldsymbol{x}\left[i-d\right]\overset{a.s.}{\to}0, \quad (64)$$

where the convergence follows a similar argument as in the proof for ZF processing. Similarly, we can show that

$$\frac{1}{N_{\mathrm{rx}}}\boldsymbol{w}_k^T\boldsymbol{n}_{\mathtt{R}}\left[i\right]\overset{a.s.}{\to}0. \quad (65)$$

Substituting (62), (63), (64), and (65) into (10), we obtain (20).

B Proof of Theorem 3

B.1 Derive $R_{\mathtt{SR}_k}$

From (27), we need to compute $\mathbb{E}\left\{\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right\}$, $\mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right)$, $\mathtt{MP}_k$, $\mathtt{LI}_k$, and $\mathtt{AN}_k$.

- Compute $\mathbb{E}\left\{\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right\}$:

  Since, $\boldsymbol{W}^T = \left(\hat{\boldsymbol{G}}_{\mathtt{SR}}^H \hat{\boldsymbol{G}}_{\mathtt{SR}}\right)^{-1} \hat{\boldsymbol{G}}_{\mathtt{SR}}^H$, from (7), we have

  $$\boldsymbol{W}^T \boldsymbol{G}_{\mathtt{SR}} = \boldsymbol{W}^T \left(\hat{\boldsymbol{G}}_{\mathtt{SR}} + \boldsymbol{\mathcal{E}}_{\mathtt{SR}}\right) = \boldsymbol{I}_{N_{\mathrm{rx}}} + \boldsymbol{W}^T \boldsymbol{\mathcal{E}}_{\mathtt{SR}}. \tag{66}$$

  Therefore,

  $$\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} = 1 + \boldsymbol{w}_k^T \boldsymbol{\varepsilon}_{\mathtt{SR},k}, \tag{67}$$

  where $\boldsymbol{\varepsilon}_{\mathtt{SR},k}$ is the kth column of $\boldsymbol{\mathcal{E}}_{\mathtt{SR}}$. Since $\boldsymbol{\varepsilon}_{\mathtt{SR},k}$ and $\boldsymbol{w}_k$ are uncorrelated, and $\boldsymbol{\varepsilon}_{\mathtt{SR},k}$ is a zero-mean random variable, $\mathbb{E}\left\{\boldsymbol{w}_k^T \boldsymbol{\varepsilon}_{\mathtt{SR},k}\right\} = 0$. Thus,

  $$\mathbb{E}\left\{\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right\} = 1. \tag{68}$$

- Compute $\mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right)$:

  From (67) and (68), the variance of $\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}$ is given by

  $$\begin{aligned}
  \mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k}\right) &= \mathbb{E}\left\{\left|\boldsymbol{w}_k^T \boldsymbol{\varepsilon}_{\mathtt{SR},k}\right|^2\right\} \\
  &= \left(\beta_{\mathtt{SR},k} - \sigma_{\mathtt{SR},k}^2\right) \mathbb{E}\left\{\left\|\boldsymbol{w}_k\right\|^2\right\} \\
  &= \left(\beta_{\mathtt{SR},k} - \sigma_{\mathtt{SR},k}^2\right) \mathbb{E}\left\{\left[\left(\hat{\boldsymbol{G}}_{\mathtt{SR}}^H \hat{\boldsymbol{G}}_{\mathtt{SR}}\right)^{-1}\right]_{kk}\right\} \\
  &= \frac{\beta_{\mathtt{SR},k} - \sigma_{\mathtt{SR},k}^2}{\sigma_{\mathtt{SR},k}^2 K} \mathbb{E}\left\{\mathrm{tr}\left(\boldsymbol{X}^{-1}\right)\right\} \\
  &= \frac{\beta_{\mathtt{SR},k} - \sigma_{\mathtt{SR},k}^2}{\sigma_{\mathtt{SR},k}^2} \frac{1}{N_{\mathrm{rx}} - K}, \text{ for } N_{\mathrm{rx}} > K,
  \end{aligned} \tag{69}$$

  where $\boldsymbol{X}$ is a $K \times K$ central Wishart matrix with N_{rx} degrees of freedom and covariance matrix $\boldsymbol{I}_K$, and the last equality is obtained by using [29, Lemma 2.10].

- Compute $\mathtt{MP}_k$:

  From (66), we have that $\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},j} = \boldsymbol{w}_k^T \boldsymbol{\varepsilon}_{\mathtt{SR},j}$, for $j \neq k$. Since $\boldsymbol{w}_k$ and $\boldsymbol{\varepsilon}_{\mathtt{SR},j}$ are uncorrelated, we obtain

  $$\mathbb{E}\left\{\left|\boldsymbol{w}_k^T \boldsymbol{\varepsilon}_{\mathtt{SR},j}\right|^2\right\} = \left(\beta_{\mathtt{SR},j} - \sigma_{\mathtt{SR},j}^2\right) \mathbb{E}\left\{\left\|\boldsymbol{w}_k\right\|^2\right\} = \frac{\beta_{\mathtt{SR},j} - \sigma_{\mathtt{SR},j}^2}{\sigma_{\mathtt{SR},k}^2} \frac{1}{N_{\mathrm{rx}} - K}. \tag{70}$$

Therefore,

$$\mathtt{MP}_k = p_{\mathrm{S}} \sum_{j \neq K}^{K} \frac{\beta_{\mathtt{SR},j} - \sigma^2_{\mathtt{SR},j}}{\sigma^2_{\mathtt{SR},k}} \frac{1}{N_{\mathrm{rx}} - K}. \tag{71}$$

- Compute $\mathtt{LI}_k$:

  From (29), with ZF, the LI can be rewritten as

$$\mathtt{LI}_k = p_{\mathrm{R}}\, \mathbb{E}\left\{ \boldsymbol{w}_k^T \boldsymbol{G}_{\mathtt{RR}} \boldsymbol{A}_{\mathtt{ZF}} \boldsymbol{A}_{\mathtt{ZF}}^H \boldsymbol{G}_{\mathtt{RR}}^H \boldsymbol{w}_k^* \right\}. \tag{72}$$

  From (14), we have

$$\boldsymbol{A}_{\mathtt{ZF}} \boldsymbol{A}_{\mathtt{ZF}}^H = \alpha^2_{\mathtt{ZF}} \hat{\boldsymbol{G}}_{\mathtt{RD}}^* \left(\hat{\boldsymbol{G}}_{\mathtt{RD}}^T \hat{\boldsymbol{G}}_{\mathtt{RD}}^* \right)^{-1} \left(\hat{\boldsymbol{G}}_{\mathtt{RD}}^T \hat{\boldsymbol{G}}_{\mathtt{RD}}^* \right)^{-1} \hat{\boldsymbol{G}}_{\mathtt{RD}}^T. \tag{73}$$

  When $N_{\mathrm{tx}} \gg K$, we can use the law of large numbers to obtain the following approximation:

$$\hat{\boldsymbol{G}}_{\mathtt{RD}}^T \hat{\boldsymbol{G}}_{\mathtt{RD}}^* \approx N_{\mathrm{tx}} \hat{\boldsymbol{D}}_{\mathtt{RD}}, \tag{74}$$

  where $\hat{\boldsymbol{D}}_{\mathtt{RD}}$ is a $K \times K$ diagonal matrix whose (k,k)th element is $\left[\hat{\boldsymbol{D}}_{\mathtt{RD}}\right]_{kk} = \sigma^2_{\mathtt{RD},k}$. Therefore,

$$\boldsymbol{A}_{\mathtt{ZF}} \boldsymbol{A}_{\mathtt{ZF}}^H \approx \frac{\alpha^2_{\mathtt{ZF}}}{N^2_{\mathrm{tx}}} \hat{\boldsymbol{G}}_{\mathtt{RD}}^* \hat{\boldsymbol{D}}_{\mathtt{RD}}^{-2} \hat{\boldsymbol{G}}_{\mathtt{RD}}^T. \tag{75}$$

  Substituting (75) into (72) we obtain

$$\begin{aligned}
\mathtt{LI}_k &\approx p_{\mathrm{R}} \frac{\alpha^2_{\mathtt{ZF}}}{N^2_{\mathrm{tx}}}\, \mathbb{E}\left\{ \boldsymbol{w}_k^T \boldsymbol{G}_{\mathtt{RR}} \hat{\boldsymbol{G}}_{\mathtt{RD}}^* \hat{\boldsymbol{D}}_{\mathtt{RD}}^{-2} \hat{\boldsymbol{G}}_{\mathtt{RD}}^T \boldsymbol{G}_{\mathtt{RR}}^H \boldsymbol{w}_k^* \right\} \\
&= p_{\mathrm{R}} \frac{\alpha^2_{\mathtt{ZF}}}{N^2_{\mathrm{tx}}} \left(\sum_{j=1}^{K} \frac{1}{\sigma^2_{\mathtt{RD},j}} \right) \mathbb{E}\left\{ \boldsymbol{w}_k^T \boldsymbol{G}_{\mathtt{RR}} \boldsymbol{G}_{\mathtt{RR}}^H \boldsymbol{w}_k^* \right\} \\
&= p_{\mathrm{R}} \frac{\alpha^2_{\mathtt{ZF}} \sigma^2_{\mathtt{LI}}}{N_{\mathrm{tx}}} \left(\sum_{j=1}^{K} \frac{1}{\sigma^2_{\mathtt{RD},j}} \right) \mathbb{E}\left\{ \|\boldsymbol{w}_k\|^2 \right\} = \frac{\sigma^2_{\mathtt{LI}} p_{\mathrm{R}} \left(N_{\mathrm{tx}} - K\right)}{\sigma^2_{\mathtt{SR},k} N_{\mathrm{tx}} \left(N_{\mathrm{rx}} - K\right)}.
\end{aligned} \tag{76}$$

- Compute $\mathtt{AN}_k$:

  Similarly, we obtain

$$\mathtt{AN}_k = \frac{1}{\sigma^2_{\mathtt{SR},k}} \frac{1}{N_{\mathrm{rx}} - K}. \tag{77}$$

Substituting (68), (69), (71), (76), and (77) into (27), we obtain

$$R_{\mathtt{SR},k} \approx \log_2 \left(1 + \frac{p_{\mathrm{S}} \left(N_{\mathrm{rx}} - K \right) \sigma_{\mathtt{SR},k}^2}{p_{\mathrm{S}} \sum_{j=1}^{K} \left(\beta_{\mathtt{SR},j} - \sigma_{\mathtt{SR},j}^2 \right) + p_{\mathrm{R}} \sigma_{\mathtt{LI}}^2 \left(1 - \frac{K}{N_{\mathrm{tx}}} \right) + 1} \right). \tag{78}$$

B.2 Derive $R_{\mathtt{RD},k}$

From (31), to derive $R_{\mathtt{RD},k}$, we need to compute $\mathbb{E}\left\{ \boldsymbol{g}_{\mathtt{RD},k}^T \boldsymbol{a}_k \right\}$, $\mathbb{V}\mathrm{ar}\left(\boldsymbol{g}_{\mathtt{RD},k}^T \boldsymbol{a}_k \right)$, and $\mathbb{E}\left\{ \left| \boldsymbol{g}_{\mathtt{RD},k}^T \boldsymbol{a}_j \right|^2 \right\}$. Following the same methodology as the one used to compute $\mathbb{E}\left\{ \boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} \right\}$, $\mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} \right)$, and $\mathtt{MP}_k$, we obtain

$$\mathbb{E}\left\{ \boldsymbol{g}_{\mathtt{RD},k}^T \boldsymbol{a}_k \right\} = \alpha_{\mathtt{ZF}}, \tag{79}$$

$$\mathbb{V}\mathrm{ar}\left(\boldsymbol{g}_{\mathtt{RD},k}^T \boldsymbol{a}_k \right) = \frac{\left(\beta_{\mathtt{RD},k} - \sigma_{\mathtt{RD},k}^2 \right) \alpha_{\mathtt{ZF}}^2}{\sigma_{\mathtt{RD},k}^2 \left(N_{\mathrm{tx}} - K \right)}, \tag{80}$$

$$\mathbb{E}\left\{ \left| \boldsymbol{g}_{\mathtt{RD},k}^T \boldsymbol{a}_j \right|^2 \right\} = \frac{\left(\beta_{\mathtt{RD},k} - \sigma_{\mathtt{RD},k}^2 \right) \alpha_{\mathtt{ZF}}^2}{\sigma_{\mathtt{RD},j}^2 \left(N_{\mathrm{tx}} - K \right)}, \text{ for } j \neq k. \tag{81}$$

Substituting (79)–(81) into (31), we obtain a closed-form expression for $R_{\mathtt{RD},k}$:

$$R_{\mathtt{RD},k} = \log_2 \left(1 + \frac{N_{\mathrm{tx}} - K}{\sum_{j=1}^{K} \sigma_{\mathtt{RD},j}^{-2}} \frac{p_{\mathrm{R}}}{p_{\mathrm{R}} \left(\beta_{\mathtt{RD},k} - \sigma_{\mathtt{RD},k}^2 \right) + 1} \right). \tag{82}$$

Then, using (24), (78), and (82), we arrive at (35).

C Proof of Theorem 4

With MRC/MRT processing, $\boldsymbol{W}^T = \hat{\boldsymbol{G}}_{\mathtt{SR}}^H$ and $\boldsymbol{A} = \alpha_{\mathtt{MRT}} \hat{\boldsymbol{G}}_{\mathtt{RD}}^*$.

1. Compute $\mathbb{E}\left\{ \boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} \right\}$:
   We have

$$\boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} = \hat{\boldsymbol{g}}_{\mathtt{SR},k}^H \boldsymbol{g}_{\mathtt{SR},k} = \left\| \hat{\boldsymbol{g}}_{\mathtt{SR},k} \right\|^2 + \hat{\boldsymbol{g}}_{\mathtt{SR},k}^H \boldsymbol{\varepsilon}_{\mathtt{SR},k}. \tag{83}$$

   Therefore,

$$\mathbb{E}\left\{ \boldsymbol{w}_k^T \boldsymbol{g}_{\mathtt{SR},k} \right\} = \mathbb{E}\left\{ \left\| \hat{\boldsymbol{g}}_{\mathtt{SR},k} \right\|^2 \right\} = \sigma_{\mathtt{SR},k}^2 N_{\mathrm{rx}}. \tag{84}$$

2. Compute $\mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}\right)$:

   From (83) and (84), the variance of $\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}$ is given by

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}\right) &= \mathbb{E}\left\{\left|\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}\right|^2\right\} - \sigma_{\mathtt{SR},k}^4 N_{\mathrm{rx}}^2 \\ &= \mathbb{E}\left\{\left|\left\|\hat{\boldsymbol{g}}_{\mathtt{SR},k}\right\|^2 + \hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\boldsymbol{\varepsilon}_{\mathtt{SR},k}\right|^2\right\} - \sigma_{\mathtt{SR},k}^4 N_{\mathrm{rx}}^2 \\ &= \mathbb{E}\left\{\left\|\hat{\boldsymbol{g}}_{\mathtt{SR},k}\right\|^4\right\} + \mathbb{E}\left\{\left|\hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\boldsymbol{\varepsilon}_{\mathtt{SR},k}\right|^2\right\} - \sigma_{\mathtt{SR},k}^4 N_{\mathrm{rx}}^2. \end{aligned} \tag{85}$$

   By using [29, Lemma 2.9], we obtain

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},k}\right) &= \sigma_{\mathtt{SR},k}^4 N_{\mathrm{rx}}\left(N_{\mathrm{rx}}+1\right) + \sigma_{\mathtt{SR},k}^2\left(\beta_{\mathtt{SR},k} - \sigma_{\mathtt{SR},k}^2\right)N_{\mathrm{rx}} - \sigma_{\mathtt{SR},k}^4 N_{\mathrm{rx}}^2 \\ &= \sigma_{\mathtt{SR},k}^2\beta_{\mathtt{SR},k}N_{\mathrm{rx}}. \end{aligned} \tag{86}$$

3. Compute $\mathtt{MP}_k$:

   For $j \neq k$, we have

$$\mathbb{E}\left\{\left|\boldsymbol{w}_k^T\boldsymbol{g}_{\mathtt{SR},j}\right|^2\right\} = \mathbb{E}\left\{\left|\hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\boldsymbol{g}_{\mathtt{SR},j}\right|^2\right\} = \sigma_{\mathtt{SR},k}^2\beta_{\mathtt{SR},j}N_{\mathrm{rx}}. \tag{87}$$

   Therefore,

$$\mathtt{MP}_k = p_{\mathrm{S}}\sigma_{\mathtt{SR},k}^2 N_{\mathrm{rx}}\sum_{j\neq k}^{K}\beta_{\mathtt{SR},j}. \tag{88}$$

4. Compute $\mathtt{LI}_k$:

   Since $\hat{\boldsymbol{g}}_{\mathtt{SR},k}$, $\boldsymbol{G}_{\mathtt{RR}}$, and $\hat{\boldsymbol{G}}_{\mathtt{RD}}$ are independent, we obtain

$$\begin{aligned}\mathtt{LI}_k &= \alpha_{\mathtt{MRT}}^2 p_{\mathrm{R}}\,\mathbb{E}\left\{\hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\boldsymbol{G}_{\mathtt{RR}}\hat{\boldsymbol{G}}_{\mathtt{RD}}^*\hat{\boldsymbol{G}}_{\mathtt{RD}}^T\boldsymbol{G}_{\mathtt{RR}}^H\hat{\boldsymbol{g}}_{\mathtt{SR},k}^*\right\} \\ &= \alpha_{\mathtt{MRT}}^2 p_{\mathrm{R}}\left(\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2\right)\mathbb{E}\left\{\hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\boldsymbol{G}_{\mathtt{RR}}\boldsymbol{G}_{\mathtt{RR}}^H\hat{\boldsymbol{g}}_{\mathtt{SR},k}^*\right\} \\ &= \alpha_{\mathtt{MRT}}^2 p_{\mathrm{R}}\left(\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2\right)\sigma_{\mathtt{LI}}^2 N_{\mathrm{tx}}\,\mathbb{E}\left\{\hat{\boldsymbol{g}}_{\mathtt{SR},k}^H\hat{\boldsymbol{g}}_{\mathtt{SR},k}^*\right\} \\ &= p_{\mathrm{R}}\sigma_{\mathtt{LI}}^2\sigma_{\mathtt{SR},k}^2 N_{\mathrm{rx}}. \end{aligned} \tag{89}$$

5. Compute $\mathtt{AN}_k$:

   Similarly, we obtain

$$\mathtt{AN}_k = \sigma_{\mathtt{SR},k}^2 N_{\mathrm{rx}}. \tag{90}$$

Substituting (84), (86), (88), (89), and (90) into (27), we obtain

$$R_{\mathtt{SR},k} = \log_2 \left(1 + \frac{p_{\mathrm{S}} N_{\mathrm{rx}} \sigma^2_{\mathtt{SR},k}}{p_{\mathrm{S}} \sum_{j=1}^{K} \beta_{\mathtt{SR},j} + p_{\mathrm{R}} \sigma^2_{\mathtt{LI}} + 1} \right). \tag{91}$$

Similarly, we obtain a closed-form expression for $R_{\mathtt{RD},k}$, and then we arrive at (36).

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195 , Feb. 2014.

[4] H. A. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen, and E. G. Larsson, "Multi-pair amplify-and-forward relaying with very large antenna arrays," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2013, pp. 3228-3233.

[5] D. W. Bliss, P. A. Parker, and A. R. Margetts, "Simultaneous transmission and reception for improved wireless network performance," in *Proc. IEEE Workshop Statist. Signal Process. (SSP)*, Aug. 2007, pp. 478–482.

[6] D. W. Bliss, T. Hancock and P. Schniter, "Hardware and environmental phenomenological limits on full-duplex MIMO relay performance," in *Proc. Annual Asilomar Conf. Signals, Syst., Comput.*, Nov. 2012.

[7] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec. 2011.

[8] —, "Hybrid full-duplex/half-duplex relaying with transmit power adaptation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 3074–3085, Sep. 2011.

[9] —, "Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation," in *Proc. Annual Asilomar Conf. Signals, Syst., Comput.*, Nov. 2011, pp 1408–1412.

[10] G. Zheng, I. Krikidis, and B. Ottersten, "Full-duplex cooperative cognitive radio with transmit imperfections," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2498-2511, May 2013.

[11] E. Aryafar, M. A. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang, "MIDU: Enabling MIMO full duplex," in *Proc. ACM Int. Conf. Mobile Comput. Netw. (MobiCom)*, Aug. 2012.

[12] M. Duarte, *Full-duplex wireless: Design, implementation and characterization.* Rice University, Houston, TX: Ph.D. dissertation, 2012.

[13] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. Ramakrishnan, C. Rice, and N. Shankaranarayanan, "Design and characterization of a full-duplex multi-antenna system for WiFi networks." [Online]. Available: http://arxiv.org/abs/1210.1639.

[14] Y. Sung, J. Ahn, B. V. Nguyen and K. Kim, "Loop-interference suppression strategies using antenna selection in full-duplex MIMO relays," in *Proc. Int. Symp. Intelligent Signal Process. and Commun. Syst. (ISPACS)*, Dec. 2011.

[15] H. A. Suraweera, I. Krikidis, and C. Yuen, "Antenna selection in the full-duplex multi-antenna relay channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2013, pp. 3416–3421.

[16] W. Zhang, X. Ma, B. Gestner, and D. V. Anderson, "Designing low-complexity equalizers for wireless systems," *IEEE Comm. Mag.*, vol. 47, pp. 56-62, Jan. 2009.

[17] Z. Zhao, Z. Ding, M. Peng, W. Wang, and K. K. Leung, "A special case of multi-way relay channel: When beamforming is not applicable," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2046-2051, July 2011.

[18] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMAX and LTE-advanced mobile systems," *IEEE Commun. Mag.*, vol 47, no 10, pp. 100-105, Oct. 2009.

[19] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Sept. 2011.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[21] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436-1449, April 2013.

[22] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[23] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[24] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[25] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 276–279. Aug. 2012.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[27] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Resource allocation for cross-layer utility maximization in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2790–2809, July 2011.

[28] W. Choi and J. G. Andrews, "The capacity gain from intercell scheduling in multi-antenna systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 714–725, Feb. 2008.

[29] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

Linköping Studies in Science and Technology
Dissertations, Division of Communication Systems
Department of Electrical Engineering (ISY)
Linköping University, Sweden

Erik Axell, *Spectrum Sensing Algorithms Based on Second-Order Statistics,* Dissertation No. 1457, 2012.

Tumula V. K. Chaitanya, *HARQ Systems: Resource Allocation, Feedback Error Protection, and Bits-to-Symbol Mappings,* Dissertation No. 1526, 2013.

Johannes Lindblom, *The MISO Interference Channel as a Model for Non-Orthogonal Spectrum Sharing,* Dissertation No. 1555, 2014.

Reza Moosavi, *Improving the Efficiency of Control Signaling in Wireless Multiple Access Systems,* Dissertations, No. 1556, 2014.

Mirsad Čirkić, *Efficient MIMO Detection Methods,* Dissertations, No. 1570, 2014.