UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Felipe Augusto Pereira de Figueiredo

On Large Scale Antenna Systems and Their Applications to Machine Type Communications
-
Estudo Sobre MIMO Massivo e suas Aplicações em Comunicações de Máquina

Campinas

2019

UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Felipe Augusto Pereira de Figueiredo

On Large Scale Antenna Systems and Their Applications to Machine Type Communications

-

Estudo Sobre MIMO Massivo e suas Aplicações em Comunicações de Máquina

Thesis presented to the School of Electrical and Computer Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering, in the area of Telecommunications and Telematics.

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica, na Área de Telecomunicações e Telemática.

Orientador: Prof. Dr. Gustavo Fraidenraich

Este exemplar corresponde à versão final da tese defendida pelo aluno Felipe Augusto Pereira de Figueiredo, e orientada pelo Prof. Dr. Gustavo Fraidenraich

---

Campinas

2019

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

F469L

Figueiredo, Felipe Augusto Pereira de, 1981-
On large scale antenna systems and their applications to machine type communications / Felipe Augusto Pereira de Figueiredo. – Campinas, SP : [s.n.], 2019.

Orientador: Gustavo Fraidenraich.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Comunicação digital. 2. Sistema de comunicação sem fio. 3. Sistema de antenas. 4. Sistemas MIMO. I. Fraidenraich, Gustavo, 1975-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estudo sobre MIMO massivo e suas aplicações em comunicações de máquina
Palavras-chave em inglês:
Digital communications
Wireless communication systems
Antenna arrays
MIMO systems
Área de concentração: Telecomunicações e Telemática
Titulação: Doutor em Engenharia Elétrica
Banca examinadora:
Gustavo Fraidenraich [Orientador]
Paulo Cardieri
Eduardo Rodrigues de Lima
Luciano Leonel Mendes
Cláudio Ferreira Dias
Data de defesa: 02-08-2019
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-2167-7286
- Currículo Lattes do autor: http://lattes.cnpq.br/0188611850092267

COMISSÃO JULGADORA – TESE DE DOUTORADO

Candidato: Felipe Augusto Pereira de Figueiredo - RA: 124103

Data da Defesa: 02 de Agosto de 2019

Título da Tese: "On Large Scale Antenna Systems and Their Applications to Machine Type Communications" (Estudo Sobre MIMO Massivo e suas Aplicações em Comunicações de Máquina).

Prof. Dr. Gustavo Fraidenraich
Prof. Dr. Paulo Cardieri
Dr. Eduardo Rodrigues de Lima
Prof. Dr. Luciano Leonel Mendes
Dr. Cláudio Ferreira Dias

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

To my mother, wife and daughter.

Acknowledgements

First and foremost, I would like to thank God Almighty for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it satisfactorily. Without His blessings, this achievement would not have been possible.

I would like to extend my sincere thanks to my supervisor, Prof. Gustavo Fraidenraich, for his valuable support and supervision. It has been a great privilege and a pleasure working with him. His advice, guidance, and encouragement have been invaluable during the course of my research.

I would also like to thank Dr. Fabbryccio Akkazzha C. Machado Cardoso for the excellent collaboration in this work and for sharing his expertise and providing me with valuable guidance.

Nevertheless, none of what I have achieved in my life would have been possible without the tremendous support of my mother, Aldair Ibrantina Pereira de Figueiredo, who has always encouraged me to keep going no matter what happened.

Additionally, in a unique way, and most importantly, I would like to thank my wife, Samanta Elena Moreira de Figueiredo, for all her dedication to contributing to my life. Her support, encouragement, patience and unconditional love were undeniably the foundation on which the past seven years of my life have been built upon. She has also given me, this year, the most precious gift anyone could ever receive, my daughter, Júlia Elena Moreira de Figueiredo, which is a true jewel.

Abstract

Massive multiple-input multiple-output (MIMO) is one of the most promising technologies for the next generation of wireless communication networks because it has the potential to provide huge improvements in Spectral-efficiency (SE) and energy efficiency. This thesis comprises of some contributions to the massive MIMO field and can be divided into two parts.

In the first part, we identify issues and solutions in the key research area of massive MIMO and propose a massive MIMO-based scheme to tackle the mixed-service communication problem, where the Base Station (BS) has to serve not only human-type communication (HTC) devices but also a possible massive number of machine-type communication (MTC) devices. Our study shows that as the number of BS antennas progressively increases, the performance of sub-optimal linear detectors approaches that of optimum detectors, such as the Successive Interference Cancellation (SIC) detectors. We additionally prove that the power radiated by each one of the MTC devices can be reduced proportionally to the number of antennas deployed at the BS. This finding is of utmost importance to low-cost and power-constrained MTC devices, which are expected to work for many years without having their batteries changed.

In the second part, we study the channel estimation problem in multi-cell massive MIMO systems. The spatial multiplexing gains, promised by the massive MIMO technology, can only be achieved if accurate channel estimation is available at both the transmitter and receiver for precoding and detection operations, respectively. However, the reuse of frequencies and pilot sequences in multi-cell communications systems leads to degradation in the channel estimation performance. This degradation is mostly caused by coherent interference, which is known as pilot contamination. In this part of our study, we devise and assess the performance of a simple and effective channel estimator for multi- and Single-Carrier Frequency-Division Multiple Access (SC-FDMA) systems. We show that the proposed estimators are able to work under moderate to strong pilot contamination without previous knowledge of the channel statistics.

Keywords: Massive MIMO; Channel Estimation; Linear Detection; Pilot Contamination; Machine Type Communications.

Resumo

MIMO é uma das tecnologias mais promissoras para a próxima geração de redes de comunicação sem fio, pois tem o potencial de fornecer enormes melhorias tanto em eficiência espectral e quanto em eficiência energética. Esta tese apresenta algumas contribuições para o campo de pesquisas de MIMO massivo e pode ser dividida em duas partes.

Na primeira parte, nós identificamos problemas e soluções na área de pesquisa de MIMO massivo e propomos um esquema baseado em MIMO massivo para lidar com o problema da comunicação de serviços mistos no Uplink (UL), onde a estação radio base (BS) tem que servir não apenas dispositivos de comunicação do tipo humano (HTC) mas também um possível número massivo de dispositivos de comunicação do tipo máquina (MTC). Nosso estudo mostra que à medida que o número de antenas instaladas em uma BS aumenta progressivamente, o desempenho de detectores lineares sub-ótimos se aproxima dos detectores ótimos, como por exemplo, os detectores de Cancelamento Sucessivo de Interferência (SIC). Além disso, provamos que a energia irradiada por cada um dos dispositivos MTC pode ser reduzida proporcionalmente ao número de antenas implantadas na BS. Essa conclusão é de extrema importância para dispositivos MTC de baixo custo e com restrição de energia, que devem funcionar por muitos anos sem que suas baterias sejam trocadas.

Na segunda parte desta tese, nós estudamos o problema de estimação de canal em sistemas MIMO massivos com múltiplas células. Os ganhos de multiplexação espacial, prometidos pela tecnologia MIMO massivo, só podem ser alcançados se uma estimação de canal precisa estiver disponível tanto no transmissor quanto no receptor para operações de pré-codificação e detecção, respectivamente. No entanto, a reutilização de frequências e sequências piloto em sistemas de comunicações multicelulares leva à degradação no desempenho da estimação de canal. Essa degradação é causada principalmente por interferência coerente, conhecida como contaminação por piloto. Nesta parte de nosso estudo, propomos e avaliamos o desempenho de estimadores de canal simples e efetivos para sistemas com portadoras múltiplas e únicas. Nós mostramos que os estimadores propostos são capazes de operar sob contaminação por piloto moderada a forte sem conhecimento prévio das estatísticas do canal.

Palavras-chaves: MIMO Massivo; Estimação de Canal; Detecção Linear; Contaminação por Piloto; Comunicações de Máquina.

List of Figures

List of Tables

List of Abbreviations

3GPP	3rd Generation Partnership Project
ADAS	Advanced Driver Assistance Systems
AMP	Approximate Message Passing
AoA	Angle of Arrival
BD	Block Diagonalization
BI-GDFE	Block-iterative Generalized Decision Feedback Equalization
BER	Bit Error Rate
BS	Base Station
CP	Cylic Prefix
CSI	Channel State Information
C-RNTI	Cell-Radio Temporary Identifier
DL	Downlink
DPC	Dirty Paper Coding
eMBB	enhanced Mobile Broadband
FCSD	Fixed Complexity Sphere Decoding
FDD	Frequency Division Duplexing
FFT	Fast Fourier transform
FP	Favourable Propagation
HTC	Human Type Communications
ICI	Inter-Carrier Interference
IoT	Internet of Things
IEEE	Institute of Electrical and Electronics Engineers
IFFT	Inverse Fast Fourier transform
ISI	Inter-Symbol Interference
ITU-R	International Telecommunication Union-Radiocommunication
LAS	Likelihood Ascent Search
LDPC	Low-Density Parity-Check
LS	Least Squares
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
LOS	Line of Sight
MFB	Matched Filter Bound
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood

MMSE	Minimum Mean Square Error
MU	Multi User
MMSE-SIC	MMSE with Successive Interference Cancellation
MRC	Maximum Ratio Combining
MSE	Mean Square Error
MTC	Machine Type Communications
mMTC	massive Machine Type Communications
MVUE	Minimum Variance Unbiased Estimator
MU-MIMO	Multi User MIMO
OFDM	Orthogonal Frequency-Division Multiplex
PAPR	High Peak to Average Power Ratio
PNSCH	Physical Narrowband Shared Channel
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QAM	Quadrature Amplitude Modulation
OOBE	Out-of-Band Emission
QPSK	Quadrature Phase-Shift Keying
RRC	Radio Resource Control
SAGE	Space-Alternating Generalized Expectation
SIB	System Information Block
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference plus Noise
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
Tx-SNR	Transmit-SNR
TDD	Time Division Duplex
THP	Tomlinson-Harashima Precoding
TS	Tabu Search
UE	User Equipment
UL	Uplink
URLLC	Ultra-Reliable Low-Latency Communications
VP	Vector Perturbation
ZC	Zadoff-Chu
ZF	Zero-Forcing
MAPE	Mean Absolute Percentage Error
PDF	Probability Density Function

CDF	Cumulative Distribution Function
SIR	Signal-to-Interference Ratio
PE	Polynomial Expansion
RAN	Radio Access Network
WSN	Wireless Sensor Networks
5G	Fifth Generation
4G	Fourth Generation
C-V2X	Vehicle-to-Everything
IMT	International Mobile Communications
MRT	Maximum Ratio-Transmission
MF	Matched Filtering
PHY	Physical-layer
i.i.d.	Independent and Identically Distributed
Tx	Transmit
SC-FDMA	Single-Carrier Frequency-Division Multiple Access
SD	Sphere Decoding
SE	Spectral-efficiency

List of Symbols

$(.)^H$	Hermitian Operator (or conjugate-transpose operator)
β_{ilk}	Large-scale fading coefficients
$\cos^{-1}(.)$	Arc-cosine of a value
Δ_f	Subcarrier Frequency Spacing
$\Gamma(.)$	Gamma function
λ	Wavelength [m]
$\|.\|_F^2$	Frobenius norm
$\mathbb{C}^{M\times K}$	Complex matrix with M rows and K columns
$\mathbb{E}\left[.\right]$	Expectation Operator
$\mathbb{P}\{.\}$	Probability of a random variable
$\mathcal{CN}$	Circularly-symmetric Gaussian distribution
$\mathcal{O}(.)$	The big-$\mathcal{O}$ notation
$\Im\{.\}$	Imaginary part of a complex number
$\Re\{.\}$	Real part of a complex number
$\overset{a}{\approx}$	Equality in asymptotic sense
Φ	Matrix containing the pilots of K devices
Ψ_P	Number of paths of the wireless channel
τ_p	Number of OFDM symbols used for Pilot Transmission
τ_u	Number of OFDM symbols used for Data Transmission
$\mathbf{0}_N$	$N \times 1$ zero vector
$\mathbf{I}_K$	$K \times K$ identity matrix
$\mathrm{Cov}\left[.\right]$	Covariance Operator
$\mathrm{var}(.)$	Variance Operator
$\mathrm{Tr}(.)$	Trace Matrix Operator
$B(.,.)$	Beta function
B_c	Coherence Bandwidth
h_{ilkm}	Small-scale fading coefficients
L	Number of cells in a multi-cell system
N	Length of the pilot sequences
N_{slot}	Number of OFDM symbols in a 0.5 ms slot
N_{smooth}	Frequency Smoothness Interval
N_s	Number of subcarriers in a OFDM symbol
T_{slot}	Slot duration
T_c	Coherence time

T_g	Duration of the Cyclic Prefix
T_s	OFDM symbol interval
T_u	Useful OFDM symbol duration
ϕ_k	Pilot sequence of the k-th device

Contents

1 Introduction

The foreseen demand increase in data rate has triggered a research race for discovering new ways to increase the SE of the next-generation of mobile and wireless networks [1]. The report in [1] predicts that data rates will easily reach peaks of 10 Gbps, reaching a staggering 49 Exabytes of data transfer per month. Additionally, the next-generation of networks is also expected to serve a higher number of devices, including HTC devices and MTC devices. One of the performance requirements defined by the International Mobile Communications (IMT) requires a connection density of 1×10^6 devices/km^2 for a network to be considered Fifth Generation (5G) [2]. According to [3], the number of connected devices is forecast to be nearly 30 billion by 2023, where around 20 billion are forecast to be MTC devices. Therefore, new technologies are required to meet these demands. It is clear that in order to improve the throughput, some new technologies that can use wider transmission bandwidths or increase the SE or even both should be exploited. In this thesis, we focus our attention on techniques that improve the SE. A well-known way to increase the SE is deploying multiple antennas at the transceivers.

Massive MIMO is a promising technology for the next-generation of wireless communications networks that can provide unprecedented gains in SE. With such technology, several devices can simultaneously communicate with a BS equipped with a large number of antennas over the same time-frequency resources. Such technology can provide huge spatial multiplexing gains even if each device is equipped with a single antenna. The more antennas a BS is equipped with, the more degrees of freedom are made available and therefore, more devices can simultaneously communicate over the same time-frequency resources. Two of the main questions that arise in this case are:

(i) whether we can obtain huge spatial multiplexing gains with low complexity signal processing algorithms or not;
(ii) what is the SE that a cell can provide to individual devices sharing the same resources?

With large antenna arrays, conventional signal processing techniques (*e.g.*, maximum likelihood detection) become prohibitively complex due to their high signal dimensions. Recently, in [4], Marzetta showed that simple linear processing, *e.g.*, Maximum Ratio-Transmission (MRT) at the transmitter and MRC at the receiver, is nearly-optimal when the number of BS antennas is large compared to the number of served devices. The main benefits of massive MIMO systems are:

(i) improved data rate and communication reliability: with a BS equipped with M antennas serving K single-antenna devices, the system is able to achieve a diversity of order M and a multiplexing gain of $\min(M, K)$,

(ii) simple signal processing: with an increasing number of BS antennas, the effect of thermal noise and small scale fading is averaged out, rendering, for example, the simple MRT/MRC optimum in such case,

(iii) power efficiency: in the UL, coherent combining achieves very high array gains which allow for substantial reduction in the transmit power of each device, on the Downlink (DL), the BS can focus its energy into the spatial directions where the devices are located.

Inspired by the above discussion, this thesis studies (i) the feasibility of employing massive MIMO technology to the challenge of serving a massive number of MTC devices over the same time-frequency resources and (ii) the channel estimation problem, which is of utmost importance for massive MIMO systems to provide significant improvements in SE. In our study, we consider performance bounds for the UL of massive MIMO systems under practical constraints such as low complexity signal processing, imperfect CSI, and inter-cell interference.

1.1 Summary of Contributions and Thesis Outline

This thesis comprises of our research results in the field of large scale antenna arrays, also known as massive MIMO, tackling aspects related to channel estimation and to the scalability of the next-generation mobile networks, where hundreds to hundred thousands of low-cost MTC devices will be served by a sole BS. The present thesis can be divided into two main problems. The first one regards the application of massive MIMO to the UL mixed-service communication problem, where a BS is expected to serve both HTC devices and a great number of MTC devices. The second problem studied in this thesis is the channel estimation in multi-cell multi-user systems where the BS is equipped with a huge number of antennas. We evaluate the effect that the phenomenon known as *pilot-contamination* has on the performance of such systems. The remainder of this thesis is structured as follows:

Chapter 2: In this chapter, we identify issues and possible solutions in the key area of large-scale antenna systems, also known as MIMO systems. Additionally, we propose the use of Massive MIMO technology as a means to tackle the UL mixed-service communication problem, where a BS has to serve not only HTC devices but also a possible massive number of MTC devices. Under the assumption of an available PNSCH, devised to exclusively consume data traffic from MTC devices, the capacity of the MTC network and, in turn, that of the whole system, can be increased by letting MTC devices share the same time-frequency physical resource blocks. We study the feasibility of applying sub-optimal linear detection to the problem of detecting a large number of MTC devices sharing the same time-frequency resources at the UL of a BS equipped with a large number of antennas, M. In this chapter, we assume perfect channel knowledge, *i.e.*, the case where the BS has perfect knowledge of the channels and therefore, it doesn't need to estimate them. Our intention with this assumption is to assess the best possible case. In our study, we derive the achievable lower-bound rates for the studied sub-optimal linear detectors and show that the transmitted power of each MTC device can be reduced proportionally to M, which is a very important result for power-constrained MTC devices running on batteries. Our simulation results suggest that, as M is made progressively larger, the performance of sub-optimal linear detection methods approach the matched filter bound, also known as perfect interference-cancellation bound.

Chapter 3: This chapter extends the evaluation of the feasibility of applying Massive MIMO to tackle the UL mixed-service communication problem presented in the previous chapter. Differently from the previous chapter, in this one, we assume imperfect channel knowledge, where the BS has to estimate the channels based on pilot sequences sent by the MTC devices. All the discussion presented in this chapter is based on the imperfect channel knowledge assumption. Here we also assume the availability of the PNSCH, with which the capacity of the whole network is increased by letting groups of MTC devices share the same time-frequency physical resource blocks. Following the research line where only imperfect channel knowledge is available, in this chapter, we (i) study the possibility of employing sub-optimal linear detectors to the problem of detecting signals from a great number of MTC devices, (ii) derive lower bounds on the achievable rate for each one of the studied linear detectors and (iii) present a simple and practical channel estimator that works without previous knowledge of the large-scale channel coefficients and noise power. Our simulation results show that (i) the use of large antenna arrays can dramatically improve the SE of a BS, allowing it to serve a great number of MTC devices over the same time-frequency resources, (ii) when the number of antennas deployed at the BS grows without bound, the transmit power of each MTC device can be made inversely proportional to $\sqrt{M}$, which is of utmost importance to low-cost and power-constrained MTC devices that are expected to work for many years without having their batteries changed, (iii) the proposed channel estimator performs asymptotically, as well as the MMSE estimator, with respect to the number of antennas and the UL transmission power. Furthermore, the results also indicate that, as the number of antennas is made progressively larger, the performance of sub-optimal linear detection methods approaches the perfect interference-cancellation bound. The findings presented here and in Chapter 2 shed light on and motivate for entirely new research lines towards a better understanding of Massive MIMO for MTC networks.

Chapter 4: This chapter presents a simple channel estimator for multi-cell multi-user massive MIMO systems considering the presence of pilot contamination. We consider Rayleigh block flat-fading channels, *i.e.*, channels that are both statistically constant during a time interval (coherence time) and flat within a frequency band (coherence band). The proposed estimator addresses performance under moderate to strong pilot contamination without previous knowledge of the large-scale channel coefficients. This estimator performs asymptotically, as well as the MMSE estimator, with respect to the number of antennas without knowledge of the channel statistics. An approximate analytical MSE expression is also derived for the proposed estimator.

Chapter 5: In this chapter, we present a study on the distribution of the simple but yet effective channel estimator for multi-cell massive MIMO systems suffering from pilot-contamination proposed in chapter 4. We prove that the distribution of the proposed channel estimator can be accurately approximated by the circularly-symmetric complex normal distribution, when the number of antennas, M, deployed at the base station is greater than 10.

Chapter 6: In this chapter, we extend our study on channel estimators for multi-cell multi-user massive MIMO systems to the SC-FDMA transmission case. Due to the SC-FDMA transmission, we propose a multipath fading channel model, which considers the presence of pilot contamination, in order to study the channel estimation problem. To facilitate the channel estimation in the multipath scenario, we propose an UL training scheme that employs Zadoff-Chu (ZC) sequences. We present a simple and effective channel estimator for multipath multi-cell massive MIMO TDD systems with pilot contamination. Differently from the MMSE channel estimator, which needs previous knowledge of the channel statistics, the proposed channel estimator works under moderate to strong pilot contamination without previous knowledge of the large-scale fading coefficients and noise power. Additionally, we derive and assess an approximate analytical MSE expression for the proposed channel estimator. We show through simulations that the proposed estimator performs asymptotically, as well as the MMSE estimator, with respect to the number of antennas and multipath coefficients.

2 Introduction to Large-Scale Antenna Systems and Massive Machine-Type Communications

2.1 Introduction

We have recently witnessed dramatic changes in the way communications systems are used. These changes are, in part, due to the big rise in on-demand data consumption over mobile and wireless networks. One issue associated with the task of accommodating such changes consists of finding solutions that can meet the diverse needs of use cases regarded as market drivers for 5G networks. The development of 5G networks has been driven by a number of use cases aimed at supporting innovative applications and services [2]. The International Telecommunication Union-Radiocommunication (ITU-R) has divided 5G network applications and services into three main categories (also referred sometimes as use cases): enhanced Mobile Broadband (eMBB); Ultra-Reliable Low-Latency Communications (URLLC); and massive Machine Type Communications (mMTC) [2, 5]. They aim at significantly improving performance, scalability and (cost/energy) efficiency of the current wireless networks such as Long Term Evolution (LTE), Long Term Evolution-Advanced (LTE-A), and LTE-A Pro. These use cases and their direct requirements will demand huge improvements in comparison with the previous generation of IMT systems [2]. A non-exhaustive list of 5G applications grouped by use case and a brief explanation about them follows next.

- **eMBB**: focus on improvements to data rate, user density, latency, capacity and coverage of the current wireless networks [6, 7]. Some applications are: high-speed mobile broadband, augmented and virtual realities (*e.g.*, gaming), smart office environments, pervasive video (*i.e.*, high-resolution video everywhere), etc.
- **URLLC**: aims at allowing devices and machines to communicate with ultra-reliability, high availability and very low latency, which make it ideal for real-time applications [8–10]. Some applications are: wireless industrial control, factory automation, remote surgery, cellular Vehicle-to-Everything (C-V2X) communications, self-driving cars, smart grids, public safety, etc.
- **mMTC**: focus on enabling machine-centered communications among devices that are massive in number, battery-driven, generate bursty traffic and have low-cost, *i.e.*, Internet of Things (IoT) devices [7, 11, 12]. This use case is intended to support

applications like: smart metering, smart cities, asset tracking, remote monitoring (*e.g.*, field and body sensors), etc.

Applications within the scope of the MTC driver range from smart cities and smart grid to critical infrastructure monitoring [13–15], and from Advanced Driver Assistance Systems (ADAS) to mobile health, which includes sports/fitness and telemedicine [16, 17]. Reliability in critical infrastructure monitoring and smart grid, for example, is often achieved only through dedicated land-line connections (*i.e.*, wired connections) [18–20]. Telemedicine makes use of telecommunications and information technology systems in order to provide remote clinical health care. It involves, for example, diagnostics realized through medical data stored in cloud servers, which requires low-latency, real-time access and high capacity servers capable of dealing with massive amounts of data, *e.g.*, computerized axial tomography and magnetic resonance imaging [21–23]. Automotive infotainment, vehicular cooperation in ADAS, and pre-crash sensing and mitigation applications also require high-speed, low-latency car-to-infrastructure and car-to-car communications [24–26].

Reliability and power consumption are of huge importance for Wireless Sensor Networks (WSN), where a few to several hundreds or even thousands of low-cost and power-constrained sensor devices (in most of the WSNs, the sensors are battery-powered) need to measure environmental conditions like temperature, noise level, air pollution levels, humidity, wind speed, etc. and reliably transmit them to a central location over harsh channel conditions [27,28]. Most of the WSN use cases require the deployment of battery-powered sensors for ten years without any maintenance, meaning that the battery is expected to last a decade without being recharged [29].

As can be noticed from the previous discussion, the requirements necessary for the implementation of next-generation wireless networks (*i.e.*, 5G) are quite diverse, even within the same market driver. Scalability is yet another issue posed by IoT, as the main assumption behind it is that hundreds to hundreds of thousands of low-cost MTC devices shall be served by a single BS [30]. Scalability issues have been mainly tackled by adopting different and sometimes complementary approaches, such as sparse signal processing techniques [31], techniques brought from duty-cycled Wireless Sensor Networks [32] and new waveforms specially designed for bursty and asynchronous data transmissions [33, 34], however, until now, the use of MIMO techniques in the context of MTC networks and the scalability issue are less understood.

The sentiment shared by most researchers nowadays is that the foreseen increase in data rate will be achieved through *combined gains* [35] provided by (i) increasing the network density, *i.e.*, the addition of more radio sites with smaller cell coverage areas to the same region (extreme network densification), which consequently improves the area SE [36], (ii) increasing spectrum availability such as the introduction of new spectrum bands like mmWaves [37, 38], (iii) improving the use of licensed, unlicensed and licensed-shared spectrum bands [39, 40] with more efficient and intelligent sharing techniques, (iv) and increasing SE of digital communications systems through advances in MIMO techniques. One of the benefits resulting directly from the powerful processing gains provided by the use of large arrays of antennas (*i.e.*, massive MIMO systems) is that the majority of the physical layer signal processing and resource allocation (*i.e.*, scheduling) issues are simplified, if not solved, which is clearly not the case for systems employing only a moderate to small number of antennas [41].

Massive MIMO has been gaining significant attention and strength as a very promising candidate to improve SE and consequently increase the channel capacity in multi-user networks. Massive MIMO is a scalable technology through which large numbers of devices can simultaneously communicate through the entire allocated spectrum, *i.e.*, thanks to its many spatial degrees of freedom, the same allocated frequency band can be reused by many users at the same time [41]. In the limit, as the number of antennas, M, deployed at the BS increases, the system processing gain also increases, *i.e.*, as M tends to infinity, the processing gain tends to infinity as well. Massive MIMO not only provides high SE in a cell, but also provides a good and uniform service to a great number of devices simultaneously [41]. A consequence of this powerful processing gain is that the effects of small-scale fading and frequency dependence disappear. In [42] it is indicated that, due to the law of large numbers, the channel becomes reliable (*i.e.*, it becomes deterministic) so that each one of the subcarriers in an Orthogonal Frequency-Division Multiplex (OFDM)-based massive MIMO system considerably experiences the same channel gain. This phenomenon is known as *channel hardening* [4]. *Channel hardening* renders frequency-domain scheduling unnecessary as all subcarriers are considered equally good, and consequently, makes most of the physical layer control signaling no longer needed [43]. Additionally, the adoption of massive MIMO systems also improve frequency reuse (due to the reduced radiated power), simplifies power control (power control coefficients depend only on the large-scale fading coefficients) and decreases multi-user interference (due to the possibility of having very narrow beams as M increases) [4, 44].

This chapter is based on our previous works [45, 46]. Differently from those works, where we have only considered the Bit Error Rate (BER) analysis for perfect channel knowledge, the current work presents additional studies and results that shed light on the application of massive MIMO technology and sub-optimal linear detection to the UL mixed-service communication problem, where a BS has to serve not only HTC devices but also a possible massive number of MTC devices. The main contributions of this chapter are the following:

- A survey on existing related work that identifies main issues and candidate solutions in the key area of Massive MIMO systems. By presenting this survey, we aim at illustrating the crucial role this technology is envisioned to play in the context of 5G wireless communication systems.

- Investigation of the feasibility of applying Massive MIMO as means to address the so-called UL mixed-service communication problem, where a *single* BS simultaneously delivers services to *both* narrowband MTC devices *and* Fourth Generation (4G) wideband services to UE devices, *i.e.*, HTC devices. In this problem, the BS has to serve HTC devices and a possible massive number of MTC devices. In order to be addressed properly, the problem can be split into two sub-problems, namely, random access and data transmission problems. The first one deals with the congestion and overloading issues brought about by the massive number of devices (*i.e.*, HTC and MTC devices) trying to get access to the network during the random access phase [47,48]. The second problem tackles the challenge of accommodating the data transmissions of this huge number of devices, which might tremendously impact on the operations and quality of the provided services of a mobile network [48, 49]. With the foreseen number of connected devices raising up to tens of thousands per cell [48], a BS might easily run out of available physical resource blocks (*i.e.*, congestion due to user data packets) to accommodate the data transmissions of this huge number of devices, tremendously impacting on the operations and quality of the provided services of a mobile network. Therefore, the focus of our work is on the data transmission phase, and in this chapter, we propose a massive MIMO-based scheme where the data transmissions of a possibly huge number of MTC devices are served through the same time-frequency resources by a BS equipped with a large number of antennas. Treating MTC devices as regular UEs turns out to be an issue, as scheduling PRBs in extremely crowded networks is a nontrivial task made harder in the presence of retransmissions and intrinsic UL synchronization procedures [50–53]. Under the assumption that a PNSCH, devised to consume the data traffic generated by MTC devices, is available, the capacity of the MTC network – and, in turn, the mixed-service system's – can be increased by allowing groups of MTC devices to share the same time-frequency PRBs. The underlying idea behind the PNSCH is the exploitation of the channel's geometric scattering characteristics to spread MTC signals in the spatial domain [4]. Individual data streams conveyed by spatially spread MTC signals can be separated thanks to the powerful processing gain of the Massive MIMO setup [4], where the size of the antenna array used at the BS is at least one order of magnitude larger than the number of served MTC devices.

It is important to notice that the number of antennas is related to the number of MTC devices being actually served over the same time-frequency resources, *i.e.*, the maximum possible number of devices scheduled to transmit data during a specific interval where the number of devices in idle state is not taken into account for the specification of the number of necessary antennas.

- Based on perfect CSI, we assess and discuss the feasibility of employing simple and sub-optimal linear detection schemes (*e.g.*, MRC, ZF, and MMSE) instead of highly complex and optimal non-linear schemes at a BS equipped with a large number of antennas. We derive closed-form lower-bound expressions on the UL achievable rates for each one of the studied linear detectors for a finite number of antennas, M. Additionally, we show that even when using simple and sub-optimal linear detectors, the transmitted power of each MTC device can be reduced when the number of antennas grows. This is a very important result for power-constrained MTC devices running on batteries.

- We present several simulation results showing that (i) the BER of sub-optimal linear detection techniques approaches the perfect interference-cancellation bound [55], as the antenna array size progressively increases, (ii) the derived achievable lower-bound rates of the studied linear detectors are tight, (iii) the transmitted power of each MTC device can be reduced with M. Based on our study, we conclude that aspects like antenna array size, performance-complexity tradeoff, and balance between interference suppression and noise enhancement dictate, as expected, the performance of a given detector.

The remainder of the chapter is organized as follows. Section 2.2 provides a brief but comprehensive overview of Massive MIMO, its challenges, and solutions available at the time of this writing. Section 2.3 presents the proposed system model and provides mathematical descriptions for each one of its functional blocks: signal generation & transmission, channel model and signal detection. We also present in this section a discussion on the capacity lower and upper bounds in favourable propagation. Section 2.4 briefly discusses contemporary solutions for signal detection in Massive MIMO systems, making the case for sub-optimal linear detection methods. In section 2.5 we derive lower-bound expressions for the UL achievable rates when linear detection (*i.e.*, MRC, ZF, and MMSE) is employed with perfect CSI knowledge at the BS. The results of our simulation work are presented and discussed in Section 2.6, while Section 2.7 wraps up the chapter with concluding remarks and suggestions for future work.

Table 2.1 – Summary of Challenges & Solutions in Large-scale Multiple Antenna Systems for 5G.

Research Area	Issue	Candidate Solutions		Shortcomings and "Side Effects"	Refs
Hardware	Phase noise	Smart PHY transceiver algorithms		Efficacy yet to be demonstrated	[42]
	Power consumption	Parallel, dedicated baseband processing		Open research question	
	Proof-of-Concept	Experiments, testbeds & prototypes		Only basic capabilities demonstrated	[56]
Antenna Aspects	Antenna coupling	Multiport impedance matching RF circuits		• Diminishes bandwidth • Introduces ohmic losses • Not fully understood for large M	[57–60] [63]
	Front-back ambiguity	Dense multidimensional implementations		• Increases coupling effects • Limited to indoor environments • 3D arrays have restricted usefulness	[57, 62]
Propagation	Channel modeling	• Realistic empirical models • Sophisticated analytical models		Currently under development	[64–66, 84]
	Cluster resolution	No solution known to date		Open research question	[61]
Transceiver Design	CSI acquisition	BS sends pilots to terminals via FDD		Limited by the channel coherence time	[4, 67]
		Terminals send pilots to BS via TDD		Channel reciprocity calibration	[69–73]
				Pilot contamination problem	[74–79]
	Precoding	Linear precoding methods	• ZF • MMSE	• Computationally heavy for large M • Higher average transmit power	[61, 86]
			• MF	• Has an error floor as M increases • Higher M required for a given SINR	[61]
			• BD	Cost-effective strategies are needed	[83]
		Nonlinear precoding methods	• DPC	Extremely costly for practical deployments	[80]
			• THP • VP	Increased complexity is hard to justify	[81] [82]
	Detection	Linear filtering	• MRC	• Does not treat interference suppression	[4, 68]
			• ZF	• Does not treat noise enhancement	[55]
			• MMSE	• More complex than MRC & ZF	[61, 85]
		Iterative linear filtering	• MMSE-SIC • BI-GDFE	Computationally heavy for large M	[87, 90, 91]
		Random step search methods	• TS • LAS	More complex than MMSE-SIC	[88] [89]
		Tree-based algorithms	• SD	Complexity grows exponentially in M	[92]
			• FCSD	• $1,000$x more complex than TS • Best suitable for the $M \approx K$ case	[93]

2.2 Massive MIMO Challenges

This section discusses issues regarded as most challenging in the Massive MIMO literature. Table 2.1 lists such issues and their available solutions, each presented alongside with its side effects, *i.e.*, new issues brought about by their adoption [45, 46].

2.2.1 Impairments due to Low-cost Hardware

Large-scale multiple antenna arrays will likely be built using low-cost components to ease the introduction and leverage the penetration of the Massive MIMO technology into the market. Hardware impairments cause channel estimation errors and limit the system's achievable capacity, which theoretically should be unlimited as the number of antennas increases. This calls for solutions capable of circumventing hardware imperfections that manifest themselves as I/Q imbalance, phase noise, power-amplifier non-linearities, and quantization errors generally intrinsic to low-cost components [54]. The power-amplifier non-linearity issue is of particular concern because low-cost power amplifiers often have relaxed linearity requirements, which in turn translate into the need for reduced High Peak to Average Power Ratio (PAPR) on a per antenna element basis [42].

Savings in radiated power result from using excess antennas to simultaneously send independent data to different users, but the total power consumption should also be taken into account. In this context, an interesting research path is hardware architectures for baseband signal processing [42]. Another path of interest is experimentation, as testbeds currently available only demonstrate basic capabilities, and do not take constrained BS real estates into consideration [56]. Experimentation can also be rewarding in that experimental findings can be fed back into theory, thus rendering the development of testbeds, prototypes, and proof-of-concept experiments of utmost importance to a better understanding of the massive MIMO technology.

2.2.2 Mutual Coupling and Front-back Ambiguity

One assumption often made when modeling antenna arrays is that the separation among antenna elements is large enough to keep mutual coupling at negligible levels. This is not entirely realistic, especially in the case of a large number of antenna elements deployed as an array of constrained size and aperture. Under such practical conditions, the mutual coupling is known to substantially impact the achievable system capacity [57]. Multiport impedance matching RF circuits can cancel out such coupling effects [58], but they diminish output port bandwidth [59] and increase ohmic losses [60, Chapter 10].

Two- or three-dimensional arrays have been reported to be able to avoid front-back ambiguity. A side effect of dense implementations is that the larger the number of adjacent elements, the larger the increase of coupling effects [61]. Another fundamental shortcoming specific to 3-D settings is the incapability of extracting additional information from the elements inside the array, *i.e.*, only elements on the array surface contribute to the information capacity [62]. The optimal densities above which the performance deteriorates no matter how large is the number of elements are studied in [63] for indoor Massive MIMO BSs.

2.2.3 RF Propagation and Channel Modeling

Realistic performance assessments call for appropriate channel characterization and modeling. The Massive MIMO channel behavior, including its correlation properties and the influence of different antenna arrangements, cannot be captured otherwise. The interest raised by this issue has been (and still is) experiencing fast-paced growth, and the community has already managed to contribute towards a better understanding of the matter. In [64], channel measurements are carried out to identify and statistically model the propagation characteristics of interest. These are then fed back into an existing channel model, extending its applicability to large-scale antenna arrays.

Performance assessments should ideally be conducted using a standardized or widely accepted channel model. Some models for Massive MIMO are presented and discussed in [65]. See, *e.g.*, [66], for a discussion on modeling methods, channel categories, and their underlying properties.

2.2.4 Acquisition of Channel State Information

In conventional Frequency Division Duplexing (FDD) systems, the BS cannot harness beamforming gains until it has established a communication link with the terminals. Firstly, the BS broadcasts pilots based on which the terminals estimate their corresponding channel responses. These terminal estimates are then quantized and fed back to the BS. Such FDD approach finds limited application in Massive MIMO systems in that the amount of time-frequency resources needed for pilot transmission in the DL scales with the number of antennas, and so does the number of channel responses that must be estimated on the part of each terminal. In systems with large antenna arrays, pilot transmission time may well exceed the coherence time of the channel [4, 67].

An alternative for Massive MIMO systems is to let the terminals send pilots to the BS via TDD. The TDD approach relies on channel reciprocity, where UL channels serve as estimates of DL channels. This leads to training requirements independent of the number of antennas, M [69], and eliminates the need for CSI feedback. TDD's drawbacks are reciprocity calibration and pilot contamination: the former is a need raised by different transfer characteristics of DL/UL processing chains (*e.g.*, amplifiers, filters, local oscillators, etc. present different characteristics); the latter arises in multi-user multi-cell scenarios where the use of non-orthogonal pilot sequences causes the intended user's channel estimate to get contaminated by a linear combination of other users' channels sharing that same pilot. Reciprocity calibration and pilot decontamination are studied in [70–79], but optimal solutions are unknown to date.

2.2.5 Precoding

Multi-user interference can be mitigated at the transmit side by modifying standard single-stream beamforming techniques to support multiple streams. Precoding based on ZF or MMSE is simple for a moderate number of antennas. However, reliance on channel inversions may take its complexity and power burdens to a point hard to accommodate within very large arrays [61, 86]. Matched Filtering (MF), which comprises MRT in the DL and MRC in the UL, is known to be the simplest method [4].

Nonlinear precoding methods, such as Dirty Paper Coding (DPC) [80], Tomlinson-

Harashima Precoding (THP) [81], and Vector Perturbation (VP) [82], also have appealing features (DPC is theoretically optimal) but are either too costly for practical deployment or offer gains hard to justify in view of their increased computational complexity. Recalling that the array size required to achieve a given Signal-to-Interference plus Noise (SINR) with MF is at least two orders of magnitude larger than with ZF [61], further work on cost-effective solutions are needed, *e.g.*, as illustrated in [83] for Block Diagonalization (BD) algorithms.

2.2.6 Detection

When it comes to data stream separation in conventional systems, Maximum Likelihood (ML) detection is the optimum solution but its complexity grows exponentially with the number of streams (this makes it hard to implement in MTC networks where hundreds to thousands of devices are envisioned). This is the reason why parameter estimation and detection are key problems in Massive MIMO systems. Suboptimal linear filtering detectors with reduced computational complexity, such as MRC, ZF, and MMSE [55], offer lower costs (that do not depend on the number of streams/users and modulation order), but are not capable of achieving the full receive-diversity order of ML detection and, consequently, they do not achieve the channel sum capacity for cases where the number of streams/users is approximately equal or equal to the number of antennas [42,43,86]. This performance-complexity tradeoff led to the development of several alternative detection methods, some of them are discussed in the sequel.

The first class of interest is iterative linear filtering, which encompasses MMSE with Successive Interference Cancellation (MMSE-SIC) and Block-iterative Generalized Decision Feedback Equalization (BI-GDFE) [87]. A shortcoming common to such iterative detectors is that their reliance on repeated matrix inversions may render them computationally heavy for large array sizes. Tabu Search (TS) [88] and Likelihood Ascent Search (LAS) [89] belong to a class of matrix-inversion free detectors known as random step search detection methods. Regrettably, the performance-complexity tradeoff comes into play also here, as both TS and LAS are known to be outperformed by MMSE-SIC [61]. Additionally, MMSE-SIC is known to achieve the sum capacity of the fast-fading MIMO multiple-access channel [90,91]. The last relevant class, referred to as tree-based detection algorithms, has in Fixed Complexity Sphere Decoding (FCSD) one of its most prominent methods [92,93]. Notwithstanding the improvements of FCSD over standard Sphere Decoding (SD), the method is still $1,000$ times more complex than TS.

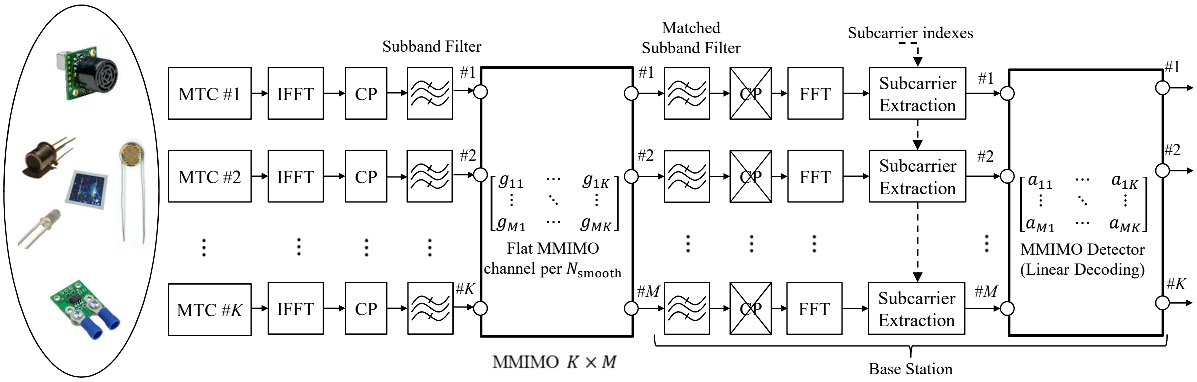

Figure 2.1 – Exemplary block diagram of a Massive MIMO UL for mixed networks, where the BS simultaneously delivers narrowband services to MTC devices and wideband services to regular UEs. The cluster of MTC devices seen at the transmit side share the same PRBs in frequency and time dimensions, while the sole BS at the receive side is equipped with an antenna array at least one order of magnitude larger than the number of served MTC devices.

2.3 System Model

This section describes the system depicted in Figure 2.1 in terms of its underlying functional blocks. In what follows, we assume that the transmitted signals of a cluster with K single-antenna MTC devices are detected by a Massive MIMO BS equipped with M receive antennas, where $M \gg K$. Single-antenna MTC devices are simple, inexpensive, power-efficient, and each device normally has low to moderate throughput.

2.3.1 Signal Generation & Transmission

Consider a PNSCH that is available and exclusively dedicated to services related to sporadic MTC traffic. The K MTC devices map data into a set of continuous PRBs in the frequency domain, with the subcarrier indexes providing the spectral position of the PNSCH at the physical layer level.

Once random access is not the focus of this work, in this chapter, we focus on the issue posed by the necessity of the BS to serve hundreds to thousands of MTC devices with a limited number of resources. However, several works in the literature deal with the problem of the random access of a large (or massive) number of devices [94–97]. Therefore, in this chapter, we assume that all MTC devices being served by a BS are already synchronized and connected to it before accessing the PNSCH, *i.e.*, all the MTC devices being served have already performed random access and attach procedures before any data is sent through the PNSCH.

The PNSCH is configured at the BS via broadcasting System Information Blocks (SIBs), just like with the Physical Random Access Channel (PRACH) used in current 4G systems (see, *e.g.* [98] and the references therein). This allows the number of

PNSCH transmission opportunities in the UL to be scheduled while taking into consideration discrepancies between the (likely different) capacities of MTC devices and regular UEs. The PNSCH time-frequency resources are semi-statically allocated by the BS, and are repeated periodically. Additionally, the System Information Block (SIB) messages can carry, for example, information about the pilot sequence length, which in turn, dictates the useful capacity of the PNSCH as it will, consequently, determine the remaining time destined to data symbols. The fraction of pilot and data symbols can be selected based on the network traffic characteristics of a cell or set of them, *i.e.*, the PNSCH configuration can be modified to increase the number of served MTC devices or the throughput of the ensemble (*i.e.*, the data rate of all MTC devices served through the same time-frequency resources) by increasing the number of OFDM symbols used for pilot or data transmission accordingly. Therefore, the pilot sequence length can be varied so that more MTC devices can be simultaneously served by the BS at the expense of smaller data capacity.

We assume the utilization of OFDM block-based transmissions where each MTC device transmits its signal (*i.e.*, allocated pilot sequence and data) by taking the Inverse Fast Fourier transform (IFFT) of the mapped information (*i.e.*, pilot and data symbols), and subsequently adding a Cylic Prefix (CP). It is important to remember that all MTC devices transmit their signals at the same time and frequency resources. We denote the OFDM symbol interval by T_s, the subcarrier spacing by Δ_f, the useful symbol duration by $T_u = 1/\Delta_f$, and the guard interval (*i.e.*, duration of the cyclic prefix) by $T_g = T_s - T_u$. As in [4], we call the reciprocal of the guard interval, when measured in subcarrier spacings, the *frequency smoothness interval*,

$$N_{\text{smooth}} = \frac{1}{T_g \Delta_f} = \frac{T_u}{T_g}, \tag{2.1}$$

where N_{smooth} represents the number of subcarriers over which the channel frequency response is considered smooth, *i.e.*, approximately constant [41].

A total of τ_p OFDM symbols are entirely used for transmitting pilot sequences. The remaining symbols, τ_u, within the same coherence interval (or coherence block) are used for data transmission. A coherence interval is a time-frequency space with duration equal to the coherence time, T_{c}, and bandwidth equal to the coherence bandwidth, B_{c}, see Figure 2.2. In general, the channel response is constant over N_{smooth} consecutive subcarriers and, therefore, the BS can estimate the channel for a total of $K_{\max} = \tau_p N_{\text{smooth}}$ terminals. We assume that a coherence interval consists of N_{smooth} subcarriers and $\tau_p + \tau_u$ OFDM symbols, *i.e.*, $N_{\text{smooth}} \times (\tau_p + \tau_u)$ subcarriers, over which the channel response can be approximated as being constant and flat-fading [44].

The modulated symbols (*i.e.*, the symbols carrying data of an MTC device) are assumed to be randomly and independently drawn from a digital modulation alphabet

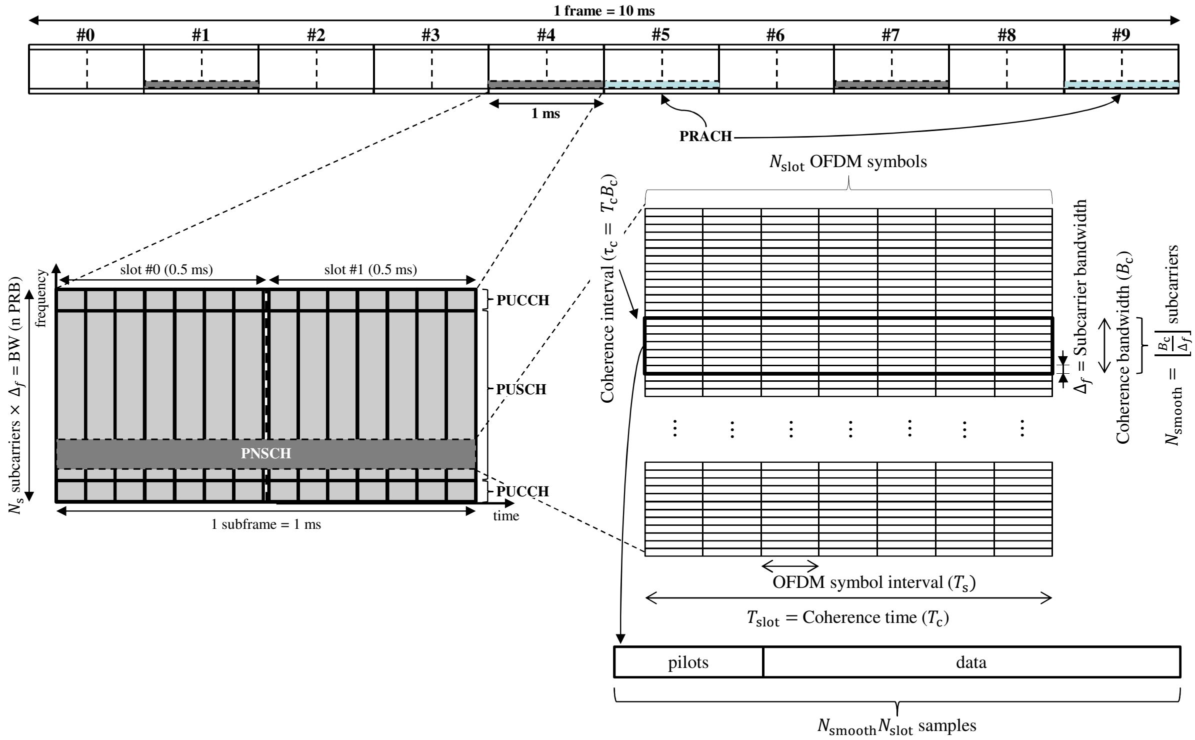

Figure 2.2 – UL frame structure with PNSCH time-frequency plane.

(*e.g.*, Quadrature Phase-Shift Keying (QPSK), 16-Quadrature Amplitude Modulation (QAM), etc.) with normalized average energy. The modulated symbols are mapped into τ_u OFDM symbols. The total number of data symbols that can be transmitted during a coherence interval is equal to $\tau_u N_{\text{smooth}}$.

Figure 2.2 depicts the UL frame structure devised for the PNSCH. As can be seen in the figure, we assume 1 ms long PNSCH transmission opportunities, however, it is important to point out that other multiple intervals of 1 ms could also be used, allowing more MTC devices and/or higher data rates. The figure also shows the time-frequency plane for one possible configuration of the PNSCH (*i.e.*, frequency position within the resource grid and time periodicity). It shows how pilots and data symbols are mapped into the time-frequency domain of a coherence interval. A 0.5 ms slot consists of N_{slot} consecutive OFDM symbols, where each of of them has N_{s} subcarriers. If we assume that $T_{\text{c}} = T_{\text{slot}}$, then, a coherence interval is composed of N_{slot} OFDM symbols and N_{smooth} consecutive subcarriers. The number of coherence intervals in a slot when $T_{\text{c}} = T_{\text{slot}}$ is given by $N_{\text{s}}/N_{\text{smooth}}$. As shown in the figure, the time-frequency plane can be divided into several coherence intervals in which each massive MIMO channel is considered time-invariant and frequency-flat.

As an example of the possible PNSCH capacity, if we consider it has 6 Physical Resource Block (PRB) allocated to it over the interval of a 1 ms long subframe (where each PRB is equal to 12 subcarriers and the subframe contains 14 OFDM symbols),

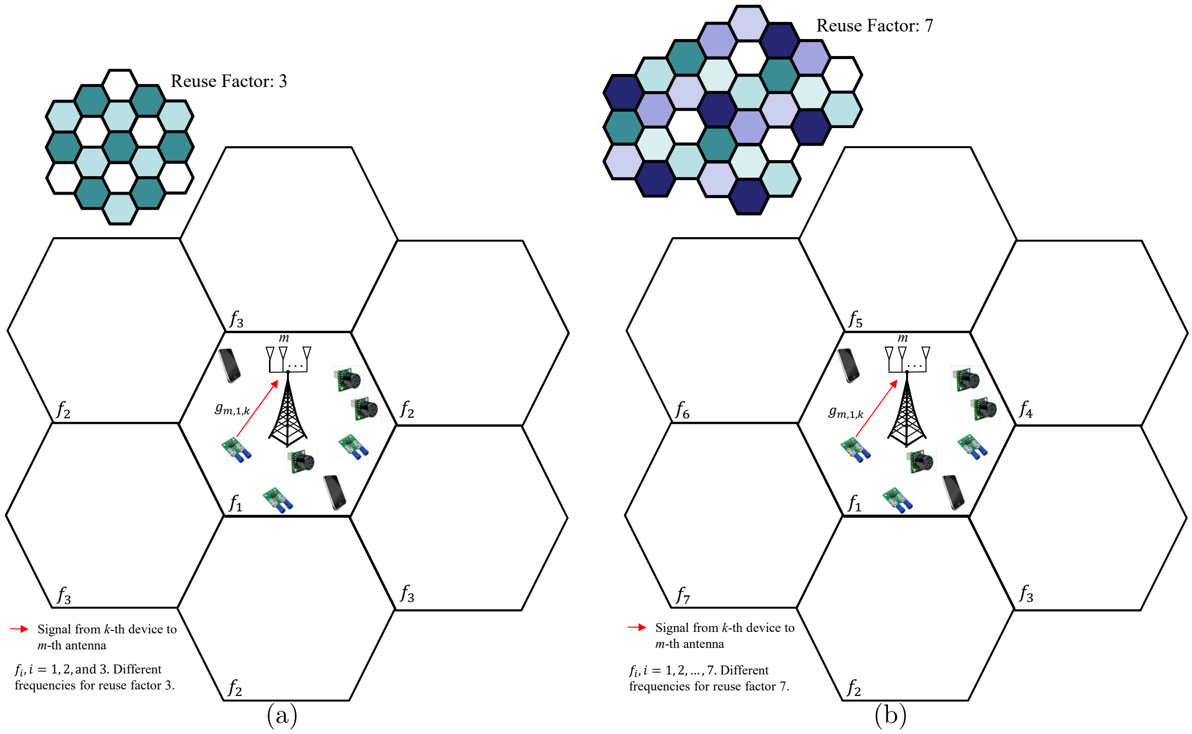

Figure 2.3 – Cell deployments with less aggressive reuse factors: (a) frequency reuse factor equal to 3 and (b) frequency reuse factor equal to 7.

a periodicity of 10 ms, meaning that MTC devices transmit for 1 ms every 10 ms, $N_{\text{smooth}} = 12$, $\tau_p = 100$, and $\tau_u = 68$, totalizing 12 subcarriers $\times$ 14 OFDM symbols $=$ 168 subcarriers/coherence interval. Therefore, for this setup, 600 different MTC devices could be served over the 6 allocated PRBs, where each one of the devices can have a data rate of 40.8 Kbits/s considering 64-QAM modulation. The throughput can be doubled (*i.e.*, 81.6 Kbits/s) if the PNSCH periodicity is decreased to 5 ms.

The PNSCH is time- and frequency-multiplexed with Physical Uplink Shared Channel (PUSCH), Physical Uplink Control Channel (PUCCH) and PRACH as depicted in Figure 2.2. Therefore, as can be seen in Figure 2.1, filters are added to both transmission and reception chains. These filters are added to the processing chains so that Out-of-Band Emission (OOBE), which are intrinsic to the OFDM waveform due to the discontinuities at its edges, do not interfere with adjacent channels, *i.e.*, PRACH, PUSCH, and PUCCH. Additionally, the filters help to mitigate Inter-Symbol Interference (ISI) and Inter-Carrier Interference (ICI) caused by asynchronous transmissions coming from random access attempts happening at the PRACH [99].

Note that the received signal is passed through a matched filter, which maximizes the SNR. The filter is applied to each time-domain OFDM symbol (*i.e.*, after IFFT and CP insertion) to mitigate the OOBE of the PNSCH transmissions. The filters should be carefully designed to (i) maintain the complex-domain orthogonality of OFDM

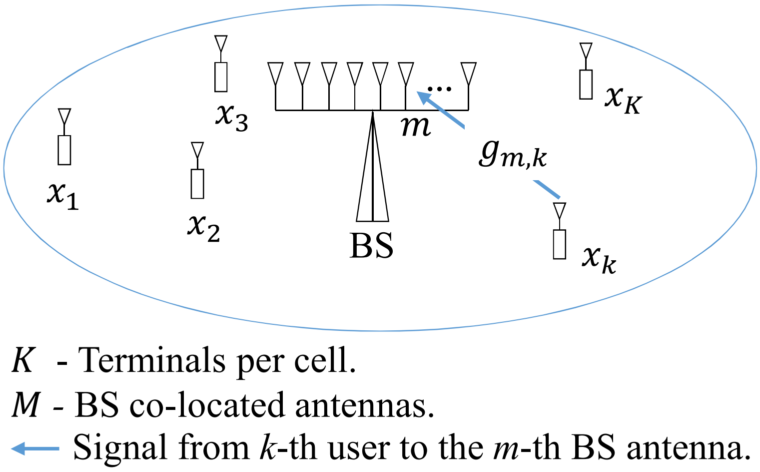

Figure 2.4 – Propagation model.

symbols, (ii) exhibit flat passband over the subcarriers in the PNSCH, (iii) have sharp transition band in order to reduce the guard-bands, and (iv) present sufficient stop-band attenuation [40, 100].

Additionally, we assume that inter-cell interference is negligible. Inter-cell interference can be heavily mitigated, and therefore, considered insignificant, if less-aggressive frequency-reuse (*e.g.*, reuses of 3, 7 or higher) is adopted [101]. Figure 2.3 depicts reuse factors equal to 3 and 7. Inter-cell interference manifests itself in two ways, namely, coherent and non-coherent interference, being the former caused by contaminating cells (*i.e.*, cells that use the same set of pilots as the home cell, causing pilot-contamination) and the latter caused by non-contaminating cells (*i.e.*, cells that do not use the same pilots as the home cell) [41]. If less aggressive frequency reuse factors are not possible or desired, then, pilot-contamination (*i.e.*, coherent interference) can be mitigated if not eliminated by making the PNSCH time-frequency resource intervals in each one of the neighbor cells different from the intervals chosen for the target cell. However, this scheme does not eliminate non-coherent interference, once the neighbor cells will be using the same frequency for other UL channels (*e.g.*, PRACH, PUCCH, PUSCH).

2.3.2 The Massive MIMO Channel and its Key Properties

Let $g_{m,k,n}$ denote the complex channel propagation coefficient from the k-th MTC device to the m-th antenna of the BS in the n-th subcarrier

$$g_{m,k,n} = h_{m,k,n}\sqrt{\beta_k}, \tag{2.2}$$

where $h_{m,k,n}$ is the complex small-scale fading coefficient, and β_k is the amplitude coefficient that accounts for geometric attenuation and shadowing, *i.e.*, large-scale fading [4]. The large-scale fading coefficients are assumed constant with respect to both subcarrier number and BS' antenna index since the geometric and shadow fading change slowly over space [4]. Therefore, between any given BS and an MTC device, there is only one large-scale fading coefficient. Additionally, these coefficients change only when an MTC device significantly changes its position in relation to the serving BS. It is generally assumed that in the radius of 10 wavelengths, the large-scale fading coefficients are approximately constant. With respect to the small-scale fading coefficients, they significantly change as the MTC device moves by a quarter of the wavelength. Therefore, it is normally assumed that the large-scale fading coefficients change about 40 times slower than the small-scale fading coefficients [102].

In general, the bandwidth occupied by an OFDM symbol is greater than the channel coherence bandwidth, B_{c}, however, on the other hand, the subcarrier bandwidth (Δ_f) is smaller than B_{c}. Therefore, as mentioned earlier, the channel can be considered flat (*i.e.*, constant) over N_{smooth} subcarriers. Additionally, we also consider that the channel is rich in scatterers and has no Line of Sight (LOS) component. Therefore, it is natural to assume that the transmitted signal within a coherence interval undergoes flat i.i.d. Rayleigh fading. The justifications for the use of this model are as follows: (i) it is approximately correct under conditions of dense scattering and (ii) it enables a comprehensive performance analysis [41].

The complex random channel responses within one coherence interval are statistically identical to the ones in any other coherence interval, irrespective of whether they are separated in time and/or frequency. Another important point is that the channel fading can be described by a stationary ergodic random process. Therefore, hereafter, our analysis is carried out by studying a single statistically representative coherence interval [41]. We assume that the channel realizations are independent between any pair of intervals, which is known as a block/interval fading assumption in the literature [4, 41]. Consequently, for notational simplicity we suppress the dependency of $g_{m,k,n}$ on the subcarrier index, n, and rewrite it as $g_{m,k}$ (see Figure 2.4).

Therefore, within a coherence interval, the $M \times K$ channel matrix $\mathbf{G}$ can be expressed as

$$\begin{aligned} \mathbf{G} &= [\mathbf{g}_1 \;\; \mathbf{g}_2 \;\; \cdots \;\; \mathbf{g}_K] = \mathbf{H}\mathbf{D}^{1/2} \\ &= \underbrace{\begin{pmatrix} h_{1,1} & \cdots & h_{1,K} \\ \vdots & \ddots & \vdots \\ h_{M,1} & \cdots & h_{M,K} \end{pmatrix}}_{\mathbf{H}} \cdot \underbrace{\begin{pmatrix} \sqrt{\beta_1} & & \\ & \ddots & \\ & & \sqrt{\beta_K} \end{pmatrix}}_{\mathbf{D}^{1/2}}, \end{aligned} \tag{2.3}$$

where the elements $g_{m,k} = h_{m,k} \cdot \sqrt{\beta_k}$ correspond to the complex channel gains from the MTC transmit antenna to the BS receive antennas where $\mathbf{g}_k \sim \mathcal{CN}(\mathbf{0}_M, \beta_k \mathbf{I}_M), \forall k$. The channel model in (2.3) is called uncorrelated Rayleigh fading or i.i.d. Rayleigh fading,

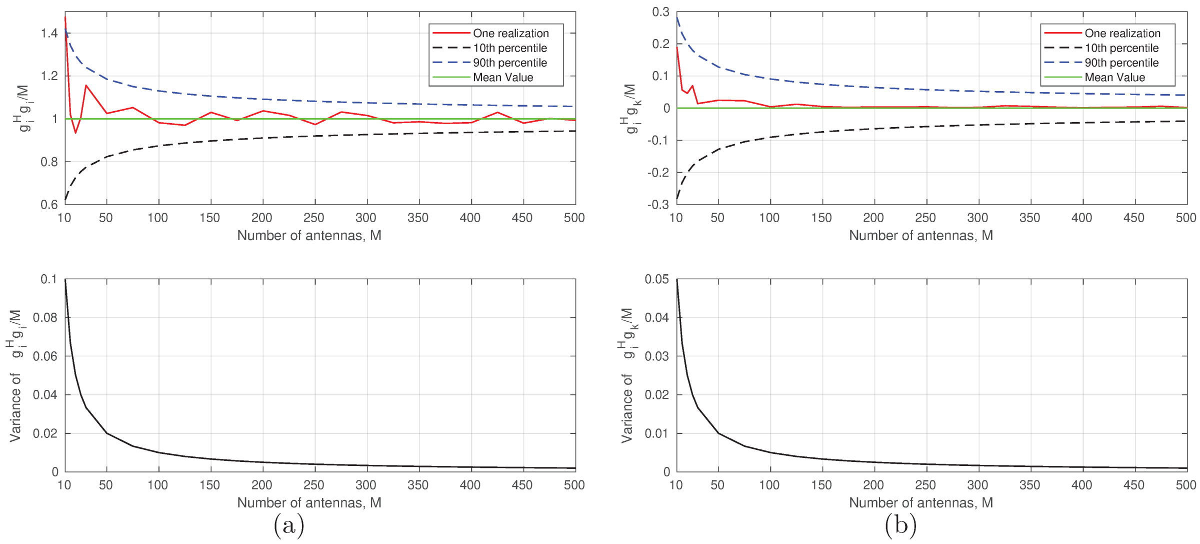

Figure 2.5 – Massive MIMO properties: (a) asymptotic channel hardening and (b) asymptotic FP

because the elements of $\mathbf{g}_k$, *i.e.*, $g_{m,k}$, are uncorrelated (and also independent) and have Rayleigh distributed magnitudes.

Small-scale fading is one of the major impairments in wireless communications [103]. This kind of fading is created by microscopic changes in the propagation environment and causes the channel gain to randomly fluctuate. The random fluctuation of the channel gain will occasionally make it very small, which in consequence causes the transmitted data to be received in error with high probability, *i.e.*, the random fluctuation makes the channel unreliable [103].

As we will discuss next, the diversity obtained by transmitting a signal over several channels with independent realizations is essential to mitigate small-scale fading and noise. Therefore, the exploitation of spatial diversity becomes very interesting, once it can be achieved by simply deploying several antennas at the receiver or transmitter.

Under the assumptions of large M and that the small-scale fading coefficients experienced by each MTC device are i.i.d. complex normal random variables with zero mean and unitary variance, the column channel vector from different MTC devices becomes asymptotically orthogonal as the number of receive antennas at the BS grows without bound [4]. On the other hand, as the number of antennas increases, the channels between the MTC devices and the BS start behaving as if they were non-fading channels, *i.e.*, as if they were almost deterministic scalar channels, after combining/precoding [104]. This phenomenon is attributed to the spatial diversity obtained from having multiple receive antennas that observe independent fading realizations, which are improbable to all be zero or nearly zero simultaneously [44].

These two key properties exhibited by Massive MIMO channels are known as *FP* and *channel hardening* [101], respectively. FP implies that noise and interference can be canceled out with simple linear detectors such as the MRC, achieving optimal performance [105]. Additionally, the FP idea provides a way to quantify the ability of a BS equipped with a large number of antennas to separate the data streams of multiple MTC devices [105]. Channel hardening makes the channel variations asymptotically reduce as the number of antennas increases, in the sense that the normalized instantaneous channel gain, $\|\mathbf{g}_k\|^2/M$, converges to the deterministic average channel gain, β_k. With channel hardening, as the small-scale fading tends to vanish, there is no need to adapt the power allocation or scheduling to its fluctuations, simplifying the resource and power allocation algorithms [106]. Moreover, channel hardening provides improved reliability due to the almost deterministic channel, which consequently results in lower latency due to a smaller number of re-transmissions. Both effects are a direct consequence of the law of large numbers, *i.e.*, they are only valid in the context of Massive MIMO when $M \to \infty$ [101]. Therefore, if the channel offers asymptotically FP and channel hardening, then we have

$$\frac{\mathbf{G}^H\mathbf{G}}{M} = \mathbf{D}^{1/2}\ \frac{\mathbf{H}^H\mathbf{H}}{M}\ \mathbf{D}^{1/2} \approx \mathbf{D}^{1/2}\mathbf{I}_K\mathbf{D}^{1/2} \to \mathbf{D},\ M \to \infty \tag{2.4}$$

where $(\cdot)^H$ denotes conjugate-transpose (*i.e.*, the Hermitian operator) operation. Equation (2.4) mathematically summarizes both phenomenons showing that when $M \to \infty$, different channels become mutually orthogonal and the channel gains tend to their respective large-scale fading coefficients. That is, as $M \to \infty$ the small-scale fading vanishes and only the large-scale fading remains, however, it can be mitigated with power control techniques [107, 108].

Figure 2.5 illustrates both the channel hardening and FP phenomenons for an M-dimensional channel $\mathbf{g}_i \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M),\ \forall i$. The figure shows the mean values, the 10% and 90% percentiles and random realizations for both phenomenons and for different numbers of antennas. As can be seen, in the channel hardening case, the normalized instantaneous channel gain $g_i^H g_i/M$ approaches its average value of 1 and the standard deviations and variance reduce as M increases. In the FP case, the inner product of the normalized channels $g_i^H/\sqrt{M}$ and $g_k/\sqrt{M}$, tends to zero and the standard deviations and variance also reduce as M increases.

We refer the interested reader to [41, 44, 109] for a detailed discussion on these phenomenons, and to [61] for experimental evidence supporting the assumption of i.i.d. Rayleigh small-scale fading coefficients with zero mean in Massive MIMO settings. In [41] the authors demonstrate that two fundamentally different channel models, namely, independent Rayleigh fading (*i.e.*, isotropic scattering) and Uniformly Random LOS, offer approximately FP and channel hardening properties. These two models represent quite extreme cases, and actually, the real channel is more likely to be something (*i.e.*, exhibit characteristics) between these two extremes. Therefore, it is reasonable to expect that the assumptions of FP and channel hardening hold for most of the practical cases. This has been experimentally confirmed by several independent measurement campaigns [42, 61, 110–113].

2.3.3 Capacity Lower and Upper Bounds in FP

The FP phenomenon does not only provide optimal performance with sub-optimal linear processing (see (2.4)) but also constitutes the most valuable scenario from the perspective of maximizing the achievable rate. In order to find lower and upper bounds for a channel offering FP we employ a capacity argument where we consider the UL direction and a fixed and deterministic channel, $\mathbf{G}$. The sum capacity for this channel is defined by [41]

$$R = \log_2 \left| \mathbf{I}_M + \rho \mathbf{G}\mathbf{G}^H \right|, \tag{2.5}$$

where ρ is the power transmitted per MTC device and we assume that the BS knows the channel $\mathbf{G}$ and that the MTC devices know their respective individual rates [41].

In order to determine the lower and upper bounds that (2.5) can assume for distinct characteristics of $\mathbf{G}$, we rewrite it as

$$\begin{aligned} R &= \log_2 \left| \mathbf{I}_M + \rho \mathbf{G}\mathbf{G}^H \right| \\ &\overset{(a)}{=} \log_2 \left| \mathbf{I}_K + \rho \mathbf{G}^H\mathbf{G} \right| \\ &\overset{(b)}{\leq} \log_2 \left(\prod_{k=1}^{K} \left[\mathbf{I}_K + \rho \mathbf{G}^H\mathbf{G} \right]_{kk} \right) \\ &= \sum_{k=1}^{K} \log_2 \left(1 + \rho \|\mathbf{g}_k\|^2 \right), \end{aligned} \tag{2.6}$$

where in (a) we used Sylvester's determinant theorem and in (b) we used the Hadamard inequality. The equality in (b) of (2.6) holds if and only if $\mathbf{G}^H\mathbf{G}$ is a diagonal matrix (*i.e.*, the channel matrix $\mathbf{G}$ has mutually orthogonal columns), which is the case when the channel exhibits FP [44]. Next, we find lower and upper bounds for (2.6) when FP is assumed and the constraint $\|\mathbf{G}\|_F^2 = \sum_{k=1}^{K} \|\mathbf{g}_k\|^2 = \sum_{k=1}^{K} \lambda_k^2 = MK$ (*i.e.*, the columns of $\mathbf{G}$ have the same norm) is applied to the channel matrix, where $\{\lambda_k\}$ are the singular values of the channel matrix $\mathbf{G}$. Notice that when we have FP, the singular values, $\{\lambda_k\}$, are equal to the channel norms, $\{\|\mathbf{g}_k\|\}$, which are in consequence equal to $\{M\beta_k\}$. Here we consider that the channel gains, $\{\beta_k\}$, are all equal to 1.

The lower bound of (2.6), given that we have FP, is achieved when $\mathbf{G}$ has its rank equal to one, which corresponds to an LOS channel. The rank of $\mathbf{G}$ is equal to the number of non-zero singular values of $\mathbf{G}$ and determines how many data streams can be simultaneously multiplexed over the channel. This is expressed by, for example, when $\|\mathbf{g}_1\|^2 = MK, \|\mathbf{g}_2\|^2 = \cdots = \|\mathbf{g}_K\|^2 = 0$, which meets the constraint $\|\mathbf{G}\|_F^2 = MK$. Consequently, the lower bound of (2.6) is given by

$$R_{\text{lower bound}} = \log_2\left(1 + \rho MK\right). \tag{2.7}$$

The upper bound of (2.6), given that we have FP, is achieved when $\mathbf{G}$ has full-rank, *i.e.*, its rank is equal to K, and $\|\mathbf{g}_k\|^2 = M, \forall k$, meaning that the channel matrix is well-conditioned [44], once $\lambda_{\min} = \lambda_{\max} = \sqrt{M}$, which are the minimum and maximum singular values of $\mathbf{G}^H\mathbf{G}$ respectively [105]. Therefore, as can be seen, a full-rank and well-conditioned channel matrix attains the highest possible capacity. Next, in order to find the upper bound for the capacity, we apply Jensen's inequality and the constraint $\|\mathbf{G}\|_F^2 = \sum_{k=1}^{K} \|\mathbf{g}_k\|^2 = MK$ to (2.6),

$$\begin{aligned} R &\leq \sum_{k=1}^{K} \log_2\left(1 + \rho\|\mathbf{g}_k\|^2\right) = K\left[\frac{1}{K}\sum_{k=1}^{K} \log_2\left(1 + \rho\|\mathbf{g}_k\|^2\right)\right] \\ &\leq K\log_2\left(1 + \frac{\rho}{K}\sum_{k=1}^{K}\|\mathbf{g}_k\|^2\right) = K\log_2\left(1 + \frac{\rho}{K}\|\mathbf{G}\|_F^2\right) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad = K\log_2\left(1 + \rho M\right). \end{aligned} \tag{2.8}$$

Therefore, the upper bound,

$$R_{\text{upper bound}} = K\log_2\left(1 + \rho M\right), \tag{2.9}$$

is achieved only when the columns of $\mathbf{G}$ are mutually orthogonal (*i.e.*, the channel offers FP), making the equality in the first line of (2.8) hold true, and when $\{\mathbf{g}_k\}$ have the same norm, making the equality in the second line of (2.8) hold true as well.

Finally, summarizing, under the constraint $\|\mathbf{G}\|_F^2 = MK$, the achievable rate for a channel exhibiting FP is bounded as

$$\log_2\left(1 + \rho MK\right) \leq R \leq K\log_2\left(1 + \rho M\right). \tag{2.10}$$

Figure 2.6 shows the capacity lower and upper bounds for several number of antennas, M, $K = 10$ MTC devices, a channel offering FP and the application of the constraint $\|\mathbf{G}\|_F^2 = MK$ to the columns of the channel matrix.

2.3.4 Signal Detection

Here we consider the scenario where the K MTC devices simultaneously transmit signals to the BS. Let x_k, where $\mathbb{E}[|x_k|^2] = 1$, $\forall k$, be the signal transmitted from the k-th device to the BS and $K \times 1$ vector, $\mathbf{x}$, the vector containing all data symbols of

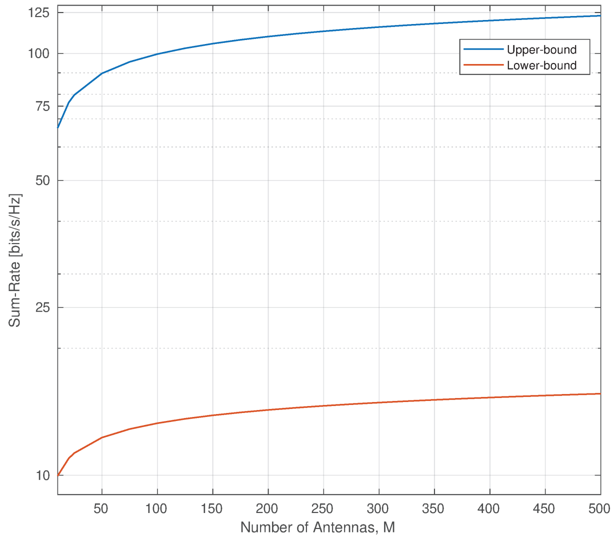

Figure 2.6 – Capacity lower and upper bounds for several number of antennas, M, when the channel offers FP.

all the K devices. Since K devices 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 devices [42, 101], which can be expressed as:

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho}\ \mathbf{G}\ \mathbf{x} + \mathbf{w} \\ &= \sqrt{\rho}\ \sum_{k=1}^{K} \mathbf{g}_k\ x_k + \mathbf{w}, \end{aligned} \tag{2.11}$$

where ρ is the average UL Transmit (Tx) power of each MTC device, $\mathbf{y} \in \mathbb{C}^{M \times 1}$, and $\mathbf{w} \in \mathbb{C}^{M \times 1}$ is a zero-mean noise vector with complex Gaussian distribution and identity covariance matrix, *i.e.*, $\mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$. As the noise variance of all elements of $\mathbf{w}$ are assumed to be equal to 1, thus, ρ can be interpreted as a normalized *transmit* SNR and consequently is dimensionless [68]. The noise variance is made unitary in order to minimize notation, but without any loss of generality. There exist M PNSCH signal versions in (2.11) for each one of the K MTC devices. Hence, the task of the BS consists of detecting K simultaneous MTC transmissions on the basis of estimates of the channel coefficients in (2.3). Therefore, detection techniques need to be employed in order to separate each of the data streams transmitted by the various devices in a Massive MIMO system.

ML multi-user detection is the optimum detection technique for the UL but it is highly complex. Its complexity grows exponentially with the number of MTC devices, K, and modulation order, making it hard to implement in our case where hundreds to thousands of MTC devices are envisioned to be served by the BS over the same time-frequency resources. To circumvent this limitation, we discuss in the next section a couple of sub-optimal alternatives with reduced computational complexity [55].

2.4 Sub-optimal Massive MIMO Detection

When it comes to separation of data streams in conventional MIMO systems (*i.e.*, where M is small), ML multi-user detection is the optimal solution but its complexity grows exponentially with the number of MTC devices. This detector works by trying all possible transmitted data vectors, $\mathbf{x}$, and selects the one that minimizes the following equation:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in X^K} \|\mathbf{y} - \sqrt{\rho}\mathbf{G}\mathbf{x}\|^2, \tag{2.12}$$

where X is the set with all possible data symbol vectors, $\mathbf{x}$. The problem (2.12) is a LS problem with a finite-alphabet constraint. The BS has to search over all possible data vectors for the one that minimizes (2.12), which as can be noticed, exponentially increases with the number of MTC devices, K (*e.g.*, for 64-QAM modulation with an alphabet of 6 symbols and 10 MTC devices the BS would have to check $6^{10} = 60466176$ possibilities). Therefore, although being the optimal solution for detection in cases where K is small, ML is a highly complex solution to be implemented in cases where hundreds to thousands of MTC devices are envisioned to be simultaneously served by a BS. This is the reason why signal detection is a key problem in Massive MIMO systems. To circumvent this limitation, in the sequel, we overview and discuss the literature on the detection subject, and justify our choices for the detectors used in the simulation work presented in Section 2.6.

2.4.1 Linear Detection Methods

Linear decoders (also known as linear detectors) work by spatially decoupling the effects of the channel by a process known as MIMO equalization. This involves multiplying $\mathbf{y}$ with a MIMO equalization matrix $\mathbf{A} \in \mathbb{C}^{M \times K}$ to get $\hat{\mathbf{x}}(\mathbf{y}) \in \mathbb{C}^{M \times 1}$ [55]. Let $\mathbf{A}$ be an $M \times K$ linear detector matrix that depends on the channel $\mathbf{G}$. By using a linear detector, the received signal vector at the BS can be separated into different data streams using $\mathbf{A}^H$ as follows

$$\mathbf{r} = \mathbf{A}^H \mathbf{y} = \mathbf{A}^H \left(\sqrt{\rho}\mathbf{G}\mathbf{x} + \mathbf{w}\right) = \sqrt{\rho}\, \mathbf{A}^H \mathbf{G}\mathbf{x} + \mathbf{A}^H \mathbf{w}. \tag{2.13}$$

where the vector $\mathbf{r}$ collects the data streams received at the BS, *i.e.*, the symbols of all K single-antenna MTC devices, and $\mathbf{A}$ is a receive matrix that depends on the specific linear detector used at the BS. As mentioned before, we consider here the case where the BS has perfect CSI, *i.e.* $\mathbf{G}$ is perfectly known at the BS, and that the channels are i.i.d. Rayleigh fading. After matched filtering, CP removal, Fast Fourier transform (FFT) processing and subcarrier extraction within each one of the OFDM symbols, as seen in Figure 2.1, a $M \times 1$ vector consisting of transmissions from all the K MTC terminals in the cell undergoes linear detection in order to retrieve the data symbols from all devices.

Inspection of (2.3) reveals that $\mathbf{D}$ is a diagonal matrix, which means we can use MRC in the UL to separate the signals from different MTC devices into different streams with asymptotic no inter-user interference as $M \to \infty$ [4]. Thereby each MTC device's transmission can be seen as the signal of a single device passing through a Single Input Single Output (SISO) channel. In the limit, this implies that MRC is optimal when the number of receive antennas is much larger than the number of transmit antennas, *i.e.* $M \gg K$, $M \to \infty$ – as can be seen from (2.4). In MRC the linear detection matrix $\mathbf{A}$ is chosen using

$$\mathbf{A}_{\text{MRC}} = \mathbf{G}, \tag{2.14}$$

where the dominant computation is due to matrix transposition. With the MRC detector, the BS aims at maximizing the received SNR of each one of the K streams, but ignoring the effect of multi-user interference. The associated complexity is of only $\mathcal{O}(MK)$ multiplications. As mentioned earlier, as M increases, MRC asymptotically becomes optimum in the sense that multi-user interference is totally removed. More specifically, the received signal is multiplied by the conjugate-transpose of the channel vector, $\mathbf{g}_k$ as follows

$$r_k = \mathbf{g}_k^H \mathbf{y} = \sqrt{\rho}\|\mathbf{g}_k\|^2 x_k + \sqrt{\rho} \sum_{i=0, i \neq k}^{K} \mathbf{g}_k^H \mathbf{g}_i x_i + \mathbf{g}_k^H \mathbf{w}. \tag{2.15}$$

One advantage of the MRC detector is that its signal processing is very simple since the BS just multiplies the received vector with the conjugate-transpose of the channel matrix $\mathbf{G}$. On the other hand, one disadvantage is that once it disregards the effect of multi-user interference it performs poorly in interference-limited scenarios.

In contrast to the MRC decoder, ZF detectors take the inter-user interference into account, but neglect the effect of noise, *i.e.*, it chooses $\mathbf{A}$ with the objective of completely eliminating inter-user interference, regardless of noise enhancement. With ZF, the multi-user interference is completely nulled out by projecting each stream onto the orthogonal complement of the inter-user interference [41]. Specifically, the ZF detector chooses $\mathbf{A}$ constrained to $\mathbf{AG} = \mathbf{I}$

$$\mathbf{A}_{\text{ZF}} = \mathbf{G}(\mathbf{G}^H \mathbf{G})^{-1}. \tag{2.16}$$

The advantages of the ZF detector are that the signal processing is simple and it works well in interference-limited scenarios. The drawback is that 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 therefore, 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 [68]. ZF exhibits a complexity of $\mathcal{O}(MK + 2MK^2 + K^3)$ [61]. With the ZF detector, the received vector, $\mathbf{y}$, is multiplied by the pseudo-inverse of the channel matrix $\mathbf{G}$, as follows

$$\mathbf{r} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{y} = \sqrt{\rho}\mathbf{x} + (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{w}. \tag{2.17}$$

For the ZF detector to work, it requires that $M \geq K$, otherwise $(\mathbf{G}^H\mathbf{G})$ is not invertible. It is noticeable that each stream of $\mathbf{r}$ in (2.17) is free of multi-user interference.

A better strategy would be to choose $\mathbf{A}$ so as to balance the signal energy lost with the increased interference. From this point of view, it is much better to accept some residual interference provided that this allows the detector to capture more of the desired signal's energy [55]. Additionally, if the channel, $\mathbf{G}$, is ill-conditioned (*i.e.*, it has a high *condition number* [44] where (2.13) has no or an infinite set of non-unique solutions), then the pseudo-inverse in (2.16) augments the noise power's, decreasing significantly ZF's performance.

One last linear detector that, together with MRC and ZF, poses complexity cost that does not depend on the number of MTC devices, K, is the MMSE. As the name suggests, the MMSE detector chooses the $\mathbf{A}$ that minimizes $e = E[\|\mathbf{A}^H\mathbf{y} - \mathbf{x}\|^2]$ without any additional constraints. The linear detection matrix for this detector is defined as

$$\mathbf{A}_{\text{MMSE}} = \mathbf{G}\left(\mathbf{G}^H\mathbf{G} + \frac{1}{\rho}\mathbf{I}\right)^{-1}. \tag{2.18}$$

In contrast to ZF, which minimizes multi-user interference but fails to treat noise, and to MRC, which minimizes noise but fails to treat interference, MMSE achieves an optimal balance between interference suppression and noise enhancement (it maximizes the received SINR [68]) at a cost (*i.e.*, complexity) similar to ZF [61, 85]. At high SNR (*i.e.*, high ρ), the ZF detector approaches the performance of the MMSE detector, while at low SNR, the MRC detector performs as well as MMSE detector [61]. The MMSE detector exhibits computational complexity similar to that of the ZF detector when the SNR, which is represented by its inverse in the second term, $\frac{1}{\rho}\mathbf{I}$, inside the parentheses of (2.18), is known. Otherwise, the complexity of the algorithm used to estimate the SNR has also to be taken into account.

2.4.2 Discussion

Table 2.1 summarizes the solutions discussed in this section as potential candidates for data stream separation (*i.e.*, detection) in the UL of Massive MIMO systems. The shortcomings listed under iterative detection, random step search, and tree-based methods suggest that these detection classes perform really well for $M \approx K$ but are too complex to be practical in the cases where hundreds to thousands of devices are served by

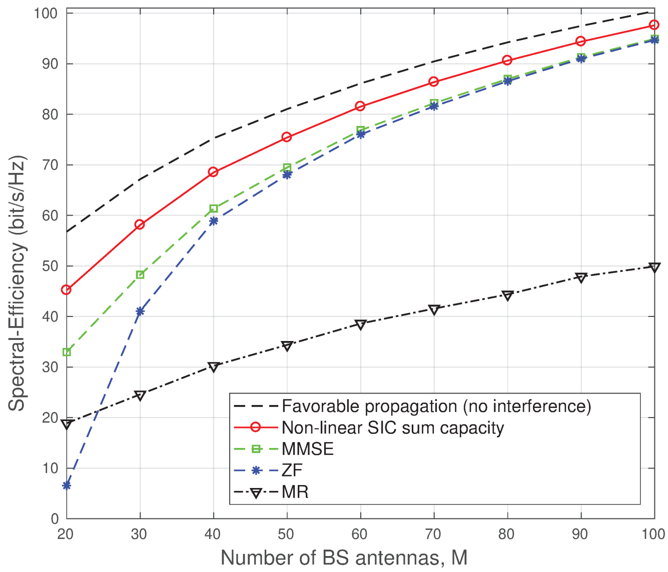

Figure 2.7 – Comparison of the sum SE for i.i.d. Rayleigh fading channels with non-linear and linear signal processing and variable number of antennas, M.

the BS and/or $M \gg K$. This indicates that more work is needed on this matter, perhaps towards turbo codes or Low-Density Parity-Check (LDPC) codes in iterative detection and decoding settings [86].

Under perfect CSI knowledge, non-linear signal detection schemes such as SIC are known to achieve the UL sum capacity [43, 90, 91]. SIC achieves the optimal performance by iteratively removing interference during the decoding process [91]. However, as mentioned earlier, this scheme requires precise CSI and a considerable number of computations. If CSI is not accurate enough, the interference removal process brings more problems than benefits [43]. On the other hand, linear detectors, such as MRC, ZF, and MMSE, can only reject interference through the application of linear transformations (*i.e.*, projections) to the received signal.

Therefore, one pertinent question that comes to our mind is how better is the performance of optimal non-linear SIC detection when compared to that of sub-optimal linear detection schemes such as MRC, ZF, and MMSE? In order to assess this question, we provide a quantitative numerical comparison in Figure 2.7. In the figure, we consider the sum capacity of a single-cell (*i.e.*, no inter-cell interference) under the availability of perfect CSI at the BS. The BS is equipped with M antennas and simultaneously serves $K = 20$ MTC devices. The channels are modeled as i.i.d. Rayleigh fading ($\mathbf{g}_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M), \forall k$) and the SNR is set to -5 dB. The figure depicts the average sum SE for different numbers of antennas, M, achieved by MRC, ZF, and MMSE sub-optimal linear detectors and the sum capacity-achieving SIC non-linear detector.

The results in Figure 2.7 show that linear processing is outperformed by non-linear SIC processing when $M \approx K$. As can be seen, MMSE detection exhibits performance that is far better than that of MRC and ZF detectors when $M \approx K$, however, it is still more than 12 bits/s/Hz lower than that presented by the SIC detector. Nevertheless, the performance gap between linear and non-linear detection schemes decreases quickly with M, once the channels become mutually orthogonal due to the FP phenomenon (*i.e.*, the channels decorrelate and then noise and interference among MTC devices vanish). As can be seen, as M increases, all the curves asymptotically approach the FP curve. Additionally, it is also important to notice that non-linear detection only significantly outperforms linear detection when $M \approx K$, while the gap becomes small when $M/K > 2$, approaching the FP curve as $M \to \infty$ [4]. These results demonstrate that a BS equipped with a large number of antennas (*i.e.*, $M \gg K$) could serve each one of the K MTC devices as if they were operating alone in the cell. It is interesting to see that the same capacity as with SIC can be achieved by using ZF and MMSE detection with the addition of a few more antennas (*e.g.*, around 10 additional antennas for this simulation setup). The addition of a few more antennas is a low price to pay for the reduced complexity of the ZF or MMSE detectors when compared to the complexity of, for example, the MMSE-SIC detector [43]. Finally, it is also important to mention that the results showed in Figure 2.7 are also valid for the DL with DPC, due to the duality explained in [114]. Therefore, as linear detectors asymptotically achieve the sum capacity as $M \to \infty$, we consider them in our simulation work.

Another important point of discussion is the comparison of the achievable capacity against the transmission power. Figure 2.8 depicts the achievable sum-rate for MRC, ZF, and MMSE detectors versus different SNR values, *i.e.*, ρ, with $M = 20$ and $K = 20$. These curves are computed by considering that the elements of $\mathbf{G}$ are i.i.d. Gaussian distributed with zero mean and unit variance. As expected, the MMSE detector performs better than ZF and MRC detectors for all SNR values. Comparing these results with those in Figure 2.7 we notice that the SE can be hugely improved by just adding antennas to the BS without the necessity to increase the transmitted power, *e.g.*, the SE for the ZF detector with $M = 20$ is 6.56 bits/s/Hz, however, just by adding 10 more antennas is grows to 41 bits/s/Hz. As we will show in the next section, the transmitted power can, in fact, be decreased as the number of antennas grows.

2.5 Achievable Rates

In this section, we present lower bounds on the achievable rate for the linear detectors discussed in the previous section when perfect CSI is assumed. Considering that a less-aggressive frequency-reuse factor (*e.g.*, reuse factor of 3, 4, 7 or higher) is adopted, we can analyze and derive the achievable rates as if the target cell was a single-

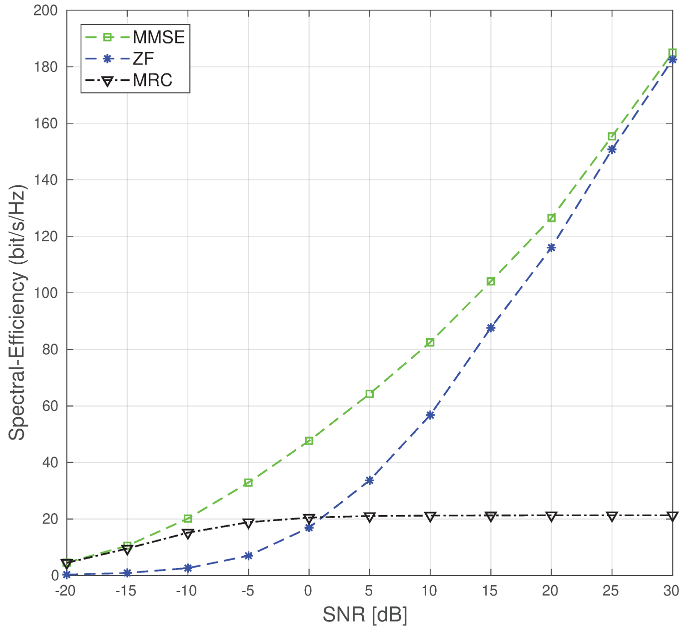

Figure 2.8 – Performance of linear receivers for different SNR values.

cell system, emphasizing the fact that inter-cell interference is inexistent or negligible and therefore, do not need to be accounted for. The rationale behind the single-cell Multi User MIMO (MU-MIMO) analysis is that its results are readily comprehended, they bound the performance of multi-cell systems and that the single-cell performance can be actually achieved if less aggressive frequency-reuse is adopted. For the following derivations, we use standard capacity bounding techniques from the massive MIMO literature [41, 101, 115].

By using (2.13), the received signal associated with the k-th MTC device is given by

$$\begin{aligned} r_k &= \sqrt{\rho}\, \mathbf{a}_k^H \mathbf{G}\mathbf{x} + \mathbf{a}_k^H \mathbf{w} \\ &= \underbrace{\sqrt{\rho}\, \mathbf{a}_k^H \mathbf{g}_k x_k}_{\text{desired signal}} + \underbrace{\sqrt{\rho} \sum_{l=1, l\neq k}^{K} \mathbf{a}_k^H \mathbf{g}_l x_l}_{\text{intra-cell interference}} + \underbrace{\mathbf{a}_k^H \mathbf{w}}_{\text{effective noise}}, \end{aligned} \tag{2.19}$$

where $\mathbf{a}_k$, and $\mathbf{g}_k$ are the k-th columns of $\mathbf{A}$, and $\mathbf{G}$, respectively. As shown in (2.19), the last two terms in the equation are considered as intra-cell interference and effective noise respectively. Therefore, an achievable rate for UL transmission for the k-th MTC device is defined by

$$R_k = \mathbb{E}\left\{\log_2\left(1 + \mathrm{SINR}_k\right)\right\}, \tag{2.20}$$

where the SINR is given by

$$\mathrm{SINR}_k = \frac{\rho |\mathbf{a}_k^H \mathbf{g}_k|^2}{\rho \sum_{l=1, l\neq k}^{K} |\mathbf{a}_k^H \mathbf{g}_l|^2 + \|\mathbf{a}_k\|^2}. \tag{2.21}$$

By following a similar line of reasoning as in [41, 101] we obtain lower bounds on the achievable rate for the MRC, ZF, and MMSE linear detectors for the perfect CSI case.

2.5.0.1 MRC detector

The received SINR of the k-th stream for the MRC detector is defined as

$$\begin{aligned}\mathrm{SINR}_k^{\mathrm{MRC}} &= \frac{\rho\|\mathbf{g}_k\|^4}{\rho\sum_{l=1,l\neq k}^{K}|\mathbf{g}_k^H\mathbf{g}_l|^2+\|\mathbf{g}_k\|^2} \\ &\rightarrow \frac{\|\mathbf{g}_k\|^4}{\sum_{l=1,l\neq k}^{K}|\mathbf{g}_k^H\mathbf{g}_l|^2}, \text{ as } \rho\rightarrow\infty.\end{aligned} \tag{2.22}$$

Considering i.i.d. Rayleigh fading, and MRC processing, the achievable rate for the k-th MTC device is lower bounded by

$$\tilde{R}_k^{\mathrm{MRC}} = \log_2\left(1+\frac{\rho(M-1)\beta_k}{1+\rho\sum_{l=1,l\neq k}^{K}\beta_l}\right). \tag{2.23}$$

If M grows without limit and we consider that the transmit power of each MTC device can be scaled with M according to $\rho = P/M$, where P is fixed, then (2.23) becomes

$$\begin{aligned}\tilde{R}_k^{\mathrm{MRC}} &= \log_2\left(1+\frac{\frac{P}{M}(M-1)\beta_k}{1+\frac{P}{M}\sum_{l=1,l\neq k}^{K}\beta_l}\right) \\ &\rightarrow \log_2\left(1+P\beta_k\right), M\rightarrow\infty.\end{aligned} \tag{2.24}$$

This result shows that with a large number of antennas and perfect CSI available at the BS, the intra-cell interference vanishes and the transmit power of each MTC device can be reduced proportionally to $1/M$. The result in (2.24) is equal to the performance of a SISO system with transmit power P, with no small-scale fading (*i.e.*, fast-fading) and intra-cell interference. Additionally, the use of a large number of antennas at the BS increases the SE of the system K times by simultaneously serving K MTC devices over the same time-frequency resources.

Remark 1. *After analyzing* (2.23) *and* (2.24)*, we see that if we decrease the transmit power proportionally to* $1/M^{\alpha}$*, where* $\alpha > 1$*, then the SINR of the UL transmission from the k-th device will go to zero as* $M\rightarrow\infty$*. When* $\alpha < 1$ *the SINR of the UL transmission from the k-th device grows without bound as* $M\rightarrow\infty$*. This means that* $1/M$ *(i.e.,* $\alpha = 1$*) is the fastest rate at which we can decrease the transmit power of each device and still maintain a fixed rate.*

2.5.0.2 ZF detector

The received SINR of the k-th stream for the ZF detector is defined as

$$\mathrm{SINR}_k^{\mathrm{ZF}} = \frac{\rho}{\left[(\mathbf{G}^H\mathbf{G})^{-1}\right]_{kk}}. \tag{2.25}$$

Considering i.i.d. Rayleigh fading, ZF processing, and $M \geq K+1$, the achievable rate for the k-th MTC device is lower bounded by

$$\tilde{R}_k^{\mathrm{ZF}} = \log_2\left(1 + \rho(M-K)\beta_k\right). \tag{2.26}$$

If M grows without limit and we make $\rho = P/M$, then (2.26) becomes

$$\begin{aligned} \tilde{R}_k^{\mathrm{ZF}} &= \log_2\left(1 + \frac{P}{M}(M-K)\beta_k\right) \\ &\to \log_2\left(1 + P\beta_k\right), M \to \infty. \end{aligned} \tag{2.27}$$

2.5.0.3 MMSE detector

The received SINR of the k-th stream for the MMSE receiver is defined as

$$\begin{aligned} \mathrm{SINR}_k^{\mathrm{MMSE}} &= \frac{1}{\left[(\mathbf{I}_K + \rho\mathbf{G}^H\mathbf{G})^{-1}\right]_{kk}} - 1 \\ &\to \frac{\rho}{\left[(\mathbf{G}^H\mathbf{G})^{-1}\right]_{kk}}, \text{ as } \rho \to \infty. \end{aligned} \tag{2.28}$$

The second line of (2.28) shows that as the SNR increases, ZF and MMSE detectors will present the same SE (see Figure 2.8).

Considering i.i.d. Rayleigh fading, and MMSE processing, the achievable rate for the k-th MTC device is approximately lower bounded by

$$\tilde{R}_k^{\mathrm{MMSE}} = \log_2\left(1 + (\alpha_k - 1)\theta_k\right), \tag{2.29}$$

where

$$\alpha_k = \frac{(M-K+1+(K-1)\mu)^2}{M-K+1+(K-1)\kappa}, \tag{2.30}$$

$$\theta_k = \frac{M-K+1+(K-1)\kappa}{M-K+1+(K-1)\mu}\rho\beta_k, \tag{2.31}$$

where μ and κ are obtained by solving the following two equations:

$$\mu = \frac{1}{K-1} \sum_{l=1, l \neq k}^{K} \frac{1}{M\rho\beta_l \left(1 - \frac{K-1}{M} + \frac{K-1}{M}\mu\right) + 1} \tag{2.32}$$

$$\begin{aligned} \kappa \left(1 + \sum_{l=1, l \neq k}^{K} \frac{\rho\beta_l}{\left(M\rho\beta_l \left(1 - \frac{K-1}{M} + \frac{K-1}{M}\mu\right) + 1\right)^2}\right) \\ = \sum_{l=1, l \neq k}^{K} \frac{\rho\beta_l\mu + 1}{\left(M\rho\beta_l \left(1 - \frac{K-1}{M} + \frac{K-1}{M}\mu\right) + 1\right)^2} \end{aligned} \tag{2.33}$$

2.6 Simulation Results

In this section, we assess the performance of MRC, ZF, and MMSE linear detectors in terms of their BER, sum-SE and transmitted power efficiency with perfect CSI knowledge. We assume a typical hexagonal single-cell structure with a radius of 1000 meters as depicted in Figure 2.3. The MTC devices are uniformly placed at random within the cell. Additionally, we consider that no device positioned closer than 100 meters to the BS. The large-scale fading coefficients $\{\beta_k\}$ are independently generated by $\beta_k = \psi / \left(\frac{r_k}{r_0}\right)^v$, where $v = 3.8$, $10 \log_{10}(\psi) \sim \mathcal{N}(0, \sigma^2_{\text{shadow,dB}})$ with $\sigma_{\text{shadow,dB}} = 8$, $r_0 = 100$ meters and r_k is the distance of the k-th MTC device to the BS. Both, the path loss exponent, v, and the standard deviation of the log-normal shadowing, $\sigma_{\text{shadow,dB}}$, are common values for outdoor shadowed urban cellular radio environments [103, 116]. We assume $K = 10$ MTC devices for all simulations in this section.

Next, we present simulations to asses how the BER behaves when a large number of antennas and sub-optimal linear detectors are employed at the BS over a range of UL transmit power (UL Tx power), ρ, values. For the following simulation results, we assume OFDM parameters similar to LTE: (i) a symbol interval of $T_s = 500/7 \approx 71.4\mu s$, (ii) a subcarrier spacing of $\Delta_f = 15$ KHz, (iii) a useful symbol duration $T_u = 1/\Delta_f \approx 66.7\mu s$, and (iv) a cyclic prefix interval (guard interval) $T_g = T_s - T_u \approx 4.76\mu s$. The frequency smoothness interval is approximately $N_{\text{smooth}} = 14$ subcarriers. We assume a coherence time of 1 ms (equivalent to 14 OFDM symbols), of which 1 symbol is used to send UL pilots and the remaining symbols, *i.e.*, 13 symbols, are used to send UL data. Therefore, in this case, the maximum number of MTC devices, $K = \tau_p N_{\text{smooth}} = 1 \times 14 = 14$. Additionally, we consider uncoded QPSK UL data transmissions. We also consider perfect time and frequency synchronism among the multiple MTC devices and the BS.

We use the MFB as a benchmark for the BER comparisons. The MFB is also known in the literature as the perfect interference-cancellation bound [55]. As the name

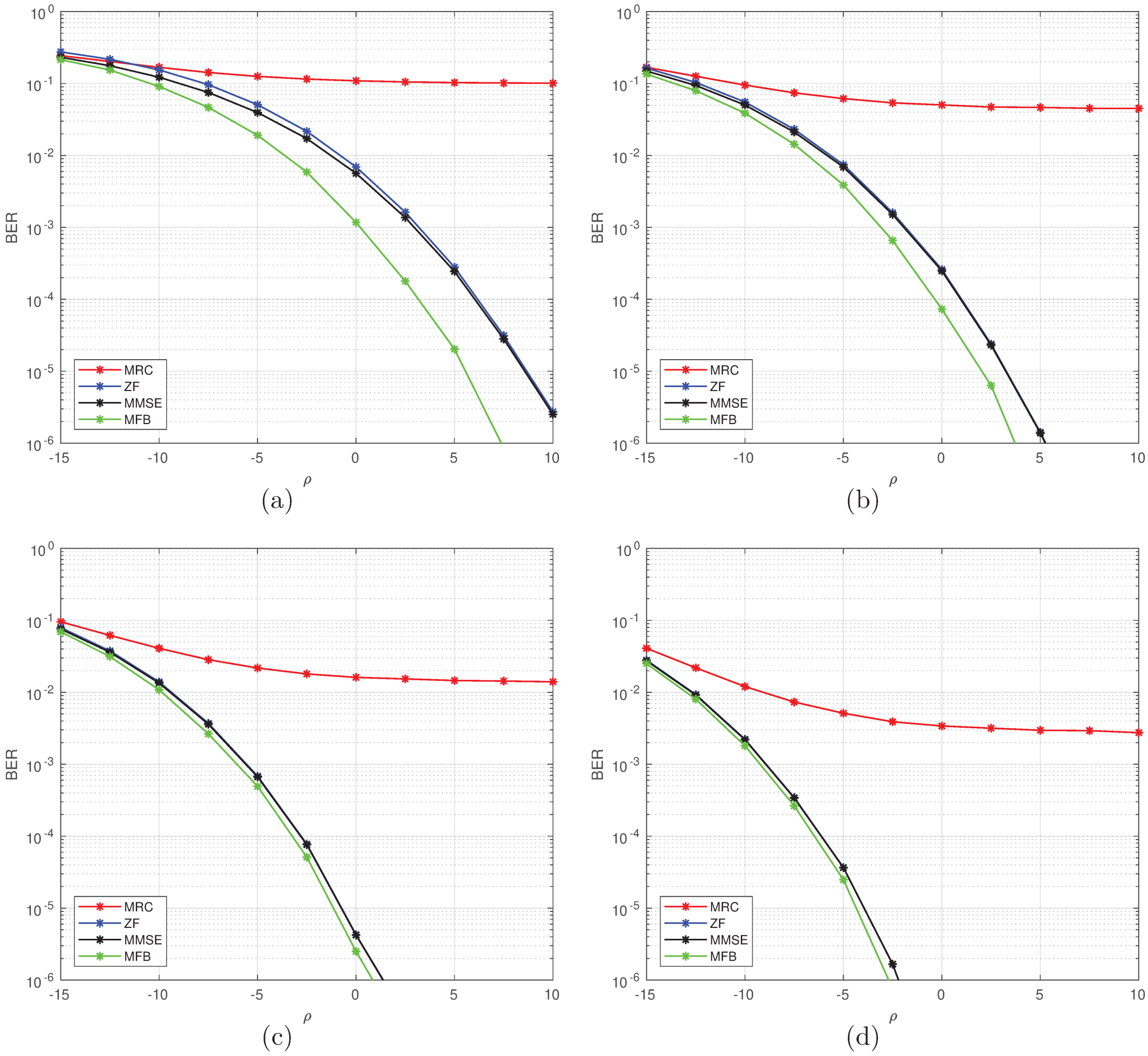

Figure 2.9 – BER performance of different linear detection methods for $K = 10$ single-antenna MTC devices over a range of different UL transmit power, ρ, values and different array sizes deployed at the BS. MFB is provided as a benchmark for comparisons. (a) $M = 20$ antennas. (b) $M = 40$ antennas. (c) $M = 80$ antennas. (d) $M = 160$ antennas.

suggests, MFB performs as the k-th user of a matched-filter receiver in the absence of no other sources of interference (*i.e.*, no cross-talk interference) [55]. Our motivation for this choice is that for $M \gg K$ both multi-user interference and noise tend to vanish (thanks to the favourable propagation phenomenon), so the performance of the $K \times M$ MIMO channel, which is assumed to be i.i.d. flat Rayleigh fading inside a coherence interval approaches the MFB as $M \to \infty$. The simulation results discussed in the sequel were averaged over 10^{10} realizations. The simulation type is Monte-Carlo with a bit error counting procedure that compares the transmitted bit vector to the received bit vector.

Figure 2.9 shows the BER of linear detectors for a fixed number of $K = 10$ MTC devices and BS array sizes in the range of $20 \leq M \leq 160$ antennas. As expected, and due to its better balance between interference suppression and noise enhancement,

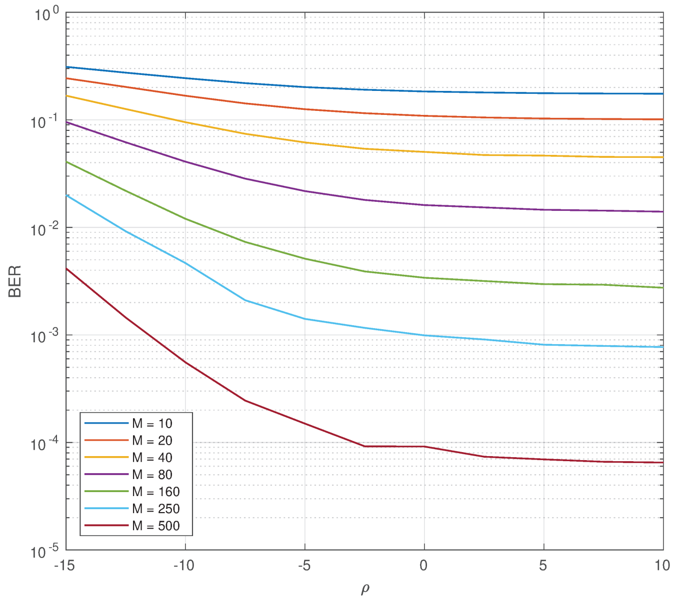

Figure 2.10 – BER performance for the MRC detector and several different number of antennas, M.

MMSE outperforms MRC and ZF in all cases studied. The performance gap inherent to MRC detection becomes evident from the results shown in this figure. The performance gap between ZF and MMSE, which is small enough to be considered negligible for $20 \leq M \leq 40$, entirely vanishes as the BS array size is grown to $M = 80$ or above. Additionally, as can also be seen, MMSE detection is the best sub-optimal linear detector among the studied detectors for cases where $M \leq K$.

Figure 2.10 shows that the performance of the MRC detector can be asymptotically increased at the expense of larger array sizes at the BS. This suggests that even the low-complexity MRC detector has the potential to approximate the MFB in the case where M can be made large enough, tending to the infinity. The main conclusion drawn from the plots is that MRC, ZF, and MMSE all approach the performance of MFB as M grows without bound, but the gap between the perfect interference-cancellation bound and ZF/MMSE decreases at a faster pace than in case of MRC.

Figure 2.11 compares the lower bounds and simulated values of the SE for MRC, ZF, and MMSE linear detectors for different number of BS antennas and perfect CSI under random large-scale fading $\{\beta_k\}$. The SE is averaged over 10^6 realizations of $\{\beta_k\}$. In this result, there are $K = 10$ MTC devices, the transmit power per terminal is $\rho = 10$ dB, the propagation channel parameters were $\sigma_{\text{shadow}} = 8$ dB, and the path loss exponent, $\gamma = 3.8$ (*i.e.*, a value for suburban areas). As can be seen, with this UL Tx power, the SE for $M = 500$ is in the order of 28 - 42 bits/s/Hz, corresponding to a SE of 2.8 - 4.2 bits/s/Hz per MTC device. These values correspond to practical values, for instance, 64-QAM modulation, where a symbol carries 6 bits, with a channel coding

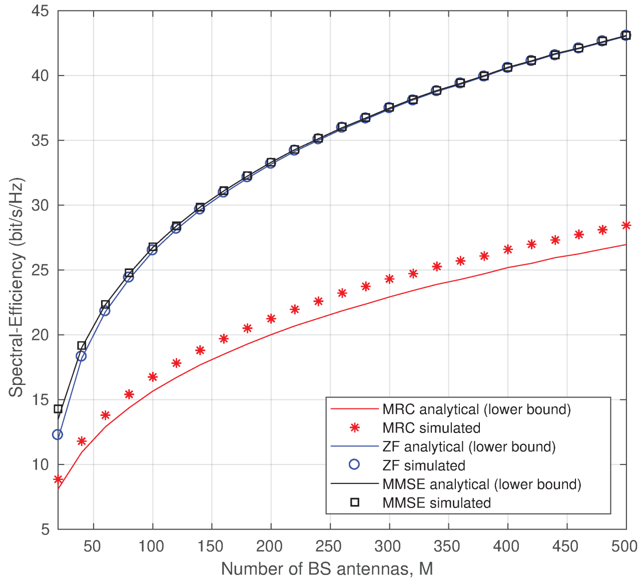

Figure 2.11 – Lower bounds and numerically evaluated values of the SE for different BS antennas for MRC, ZF, and MMSE linear detectors and perfect CSI.

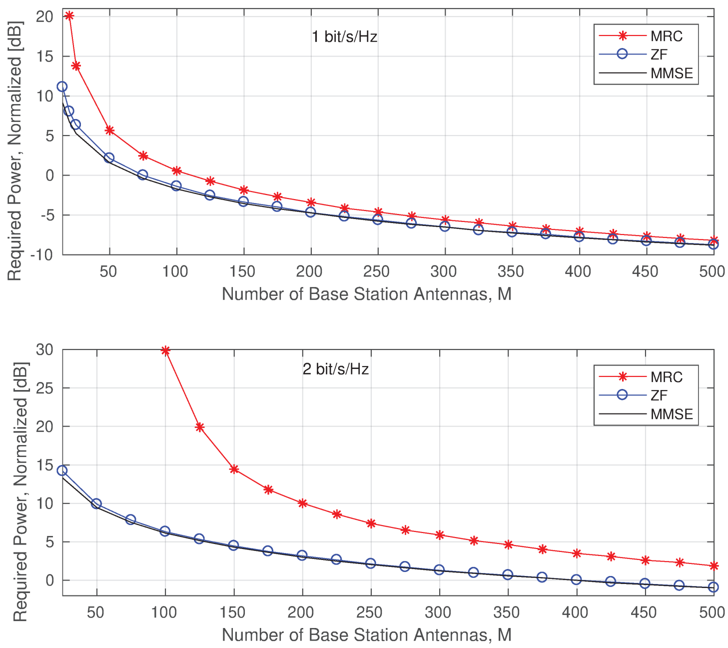

Figure 2.12 – Transmit power required to achieve 1 and 2 bits/s/Hz per MTC device for MRC, ZF, and MMSE linear receivers as a function of the number of antennas M. The number of MTC devices is set to $K = 10$, and the propagation parameters are $\sigma_{\text{shadow}} = 8$ dB and $\upsilon = 3.8$.

rate of 3/4 corresponds to a SE of 4.5 bits/s/Hz. As can be noticed from the figure, the MMSE detector always presents better performance than the MRC and ZF detectors, however, the ZF's detector performance asymptotically approaches the MMSE when M increases. Therefore, we conclude that even with simple and suboptimal linear processing algorithms such as MRC and ZF, it is possible to achieve high SEs and consequently serve more MTC devices through the same time/frequency resources.

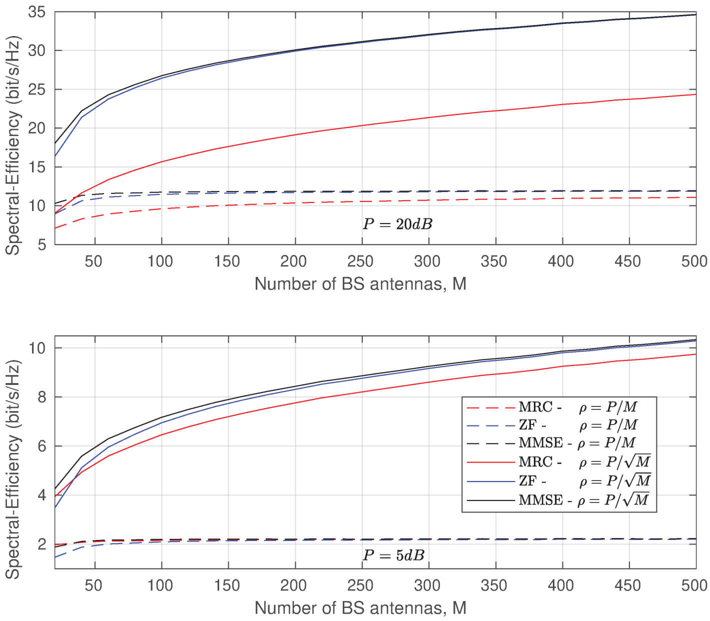

Figure 2.13 – SE versus number of antennas, M, for linear detectors: MRC, ZF, and MMSE. In these examples $K = 10$ MTC devices are served simultaneously and the transmit power is $P = 20$ dB for the upper figure and $P = 5$ dB for the lower figure and the propagation parameters are $\sigma_{\text{shadow}} = 8$ dB and $\upsilon = 3.8$.

In Figure 2.12 we show the transmit power per MTC device that is needed to reach a fixed SE. The figure shows the normalized power, ρ, required to achieve 1 and 2 bits/s/Hz per MTC device as a function of M. For the 1 bit/s/Hz case, it can be seen that by doubling the number of antennas, M, the transmit power can be cut back by approximately 3 dB. When M is large, the difference in performance between MRC and ZF (or MMSE) is less than 0.65 dB. This difference increases when the target SE is increased. For the 2 bits/s/Hz case, the cross-talk interference, *i.e.*, interference from other devices, is more significant (relative to thermal noise) and therefore, ZF and MMSE receivers perform relatively better. The difference in performance between MRC and ZF (or MMSE) for $M = 500$ antennas is approximately 2.86 dB. This characteristic of massive MIMO is very important to MTC devices, where stringent power constraints have to be respected.

In Figure 2.13 we illustrate the power-scaling laws. The figure shows the SE versus the number of antennas M for $\rho = P/M$ and $\rho = P/\sqrt{M}$ and with MRC, ZF, and MMSE linear detectors respectively. In the upper part of the figure, we set P to 20 dB. At this SNR, the SE is in the order of 10-35 bits/s/Hz, which corresponds to a SE per MTC device of 1-3.5 bits/s/Hz. As expected, with $\rho = P/M$, when the number of antennas, M, grows, the SE approaches a constant value. However, when $\rho = P/\sqrt{M}$, the SE grows without bound, *i.e.*, logarithmically fast with M, when $M \to \infty$. Next, in the lower part of the figure, we set P to 5 dB. This figure provides the same insights as the upper one. The gap between the performance of MRC and that showed by the ZF and MMSE detectors is reduced compared to the upper part of the figure. This is due to the fact that the interference caused by cross-talk, *i.e.*, the interference coming from other MTC devices, is smaller in this figure when compared to the thermal noise. These results attest that the transmit power of each MTC device can be scaled down as P/M. If we compare the MRC detector with ZF and MMSE detectors in the upper part of the figure, it is possible to see that when the radiated power by MTC device is made inversely proportional to $\sqrt{M}$, the power level is not sufficiently low to make the MRC detector perform as well as the ZF and MMSE detectors. However, when the radiated power is made inversely proportional to M, then, in this case, the MRC detector performs closely to the ZF and MMSE detectors for a large number of antennas, M. Therefore, these results show that the radiated power by MTC device can be made inversely proportional to the number of BS antennas, M, with no reduction of the sum SE performance. Additionally, as the radiated power levels are reduced, the cross-talk interference caused by the use of the inferior MRC detector eventually drops below the noise level, making this simple linear detector a practical option, once it performs as well as ZF and MMSE detectors for the large M regime.

2.7 Conclusions

In this chapter, we proposed the use of a Massive MIMO technology as a means to address the UL mixed-service communication problem by offloading the MTC data traffic to an exclusively channel, which was devised to serve such devices. In our study, we considered K single-antenna MTC devices served by a BS equipped with a large number of antennas, M. Under the assumption of an available PNSCH, the capacity of the MTC network and, in turn, that of the whole system, can be increased by grouping MTC devices into clusters and letting each cluster share the same time-frequency resources. We show that the individual data streams conveyed by spatially spread MTC signals can be separated thanks to the powerful processing gain provided by the adoption of Massive MIMO technology.

As the size of the antenna array at the BS is made progressively larger, our simulation results suggest that the BER performance of sub-optimal linear detection methods approach the MFB (also known as perfect interference-cancellation bound). ZF and MMSE detectors approach the bound at a faster pace than the simpler and less complex MRC detector. Due to its better balance between interference suppression and noise enhancement, MMSE outperforms MRC and ZF in all studied cases. The gap in the performance of the ZF detector, however, is negligible for array sizes around 40 antennas, and entirely vanishes when the BS is equipped with $M \geq 80$ antennas.

We provide results showing that the adoption of massive MIMO technology increases the SE of the whole system by allowing a great number of MTC devices to be served through the same time-frequency physical resources. Additionally, we also present results demonstrating that the transmit power of the MTC devices can be made inversely proportional to the number of available antennas.

It is also important to notice that (2.4) is an approximation that only becomes an equality when $M \gg K$, $M \rightarrow \infty$ and the channel offers FP and channel hardening, which explains why the MRC detection is not optimum for the results presented here.

3 Applying Massive MIMO Systems to Machine-Type Communications

3.1 Introduction

Recent technological developments taking place in our society have been drastically changing the way we use communications systems. These changes are in their great part due to the huge (and also foreseen [1]) increase in on-demand data consumption over both wireless and mobile networks. In order to support such changes, it is mandatory to devise solutions that can meet the different requirements of use cases regarded as the market drivers for next-generation wireless networks (*e.g.*, 5G mobile and wireless networks). As listed in chapter 2, the ITU-R has defined the following three main 5G use cases: eMBB; URLLC; and mMTC [2].

The adoption of massive MIMO technology can especially help to leverage and simplifying the deployment of mMTC systems in cellular networks, which are potential candidates to accommodate the emerging MTC data traffic thanks to the existing infrastructure and wide-area coverage [117]. Massive MIMO has the potential to enable the multiplexing of a myriad of devices in the same time/frequency resources along with an extension in range due to the coherent beamforming gain inherent to this technology [41, 117].

The main contribution of this chapter is the further study of the proposed data transmission scheme employing massive MIMO technology as a way to address the UL mixed-service communication problem presented in the previous chapter. The problem is split into two sub-problems: random access and data transmission problems. Again, as in chapter 2, we focus our work on the data transmission phase.

The transmission scheme proposed in Chapter 2 and further studied here, has the potential to mitigate the congestion due to the large (and possibly massive) number of user data packets and additionally, it offers scalability, as the number of served devices can easily grow by increasing the number of deployed antennas at the BS [41]. The proposed scheme deals with data transmission, channel estimation and detection of the many data streams simultaneously transmitted by multiple MTC devices using the PNSCH's shared time-frequency resources.

We employ the ML method to find an estimator for the large-scale fading coefficients present in the MMSE channel estimator. We show that this estimator is not only unbiased but it also achieves the Crámer-Rao lower bound. The estimated large-scale fading coefficients are replaced into the MMSE channel estimator, giving rise to a new channel estimator, which asymptotically approaches the performance of the MMSE channel estimator as the number of antennas and the UL transmit power increase. Additionally, we derive closed-form and approximate expressions for the MSE of the proposed estimator. Moreover, we find lower bounds on the achievable rate for each one of the studied linear detectors. We also show that even for simple linear receivers (*i.e.*, MRC, ZF, and MMSE), the transmitted power of each MTC device can be reduced as the number of antennas, M, grows without bound, which is very beneficial for power-constrained devices running on batteries.

Differently from chapter 2, where we have considered perfect channel estimation (*i.e.*, full channel knowledge), this chapter not only deals with imperfect channel estimation, proposing and assessing the performance of a channel estimator in terms of MSE and BER, but also proposes a scheme to tackle the problem posed by the simultaneous data transmission of a large number of MTC devices connected to the base station (*i.e.*, user data packet congestion). Additionally, we also analyze the achievable rates of the studied linear detectors with the proposed data transmission scheme.

The remainder of the chapter is as follows. Section 3.2 provides a brief discussion on related works. Section 3.3 presents a study case, where the feasibility of Massive MIMO for MTC networks is investigated as a means to address the UL mixed-service communication problem. Section 3.4 outlines one possible approach to estimate the large-scale fading coefficients. Section 3.5 proposes a channel estimator that takes into account the estimation of the large-scale fading coefficients. Section 3.6 presents simulation results and discussions on the outcomes. Section 3.7 wraps up the chapter with concluding remarks.

3.2 Related Work

In the literature, there is a myriad of works proposing solutions exclusively tailored to increase the capacity of the random access channel of LTE/LTE-A networks. In those networks, the MTC devices compete for resource blocks for their data transmission using a random access scheme. The works [48, 94–97] and the vast number of papers therein mentioned, review and propose solutions to tackle the scalability issue posed by the random access of tens to hundreds of thousands of MTC devices during the random access/synchronization phase. The solutions presented in these works can accommodate from 30000 to more than 78000 MTC devices per cell with low collision probabilities. However, these works do not deal with the problem involving simultaneous data transmissions coming from a possible massive number of MTC devices during the Radio Resource Control (RRC) connected state [49].

Although there exist numerous studies on the massive random access problem, there are relatively few publications addressing the massive data transmission problem (*i.e.*, congestion due to user data packets) [118], which arises from the simultaneous data transmission of a huge number of devices during the RRC connected state (*i.e.*, during the data transmission phase).

The work presented in [119] tackles the problem of device activity detection and joint channel estimation when non-orthogonal pilot sequences are used by the devices. The authors use the Approximate Message Passing (AMP) algorithm from compressed sensing theory to exploit the sparsity in device activity detection. The work considers a grant-free multiple-access scheme and that the devices are already synchronized to the BS. The drawbacks of the proposed solution are the lack of information on how the devices stay synchronized to the BS and the analysis for multi-cell scenarios. In [120] the authors study the coexistence of HTC and MTC devices under a single-cell massive MIMO setup and assess their joint SE, however, they do not deal with channel estimation, linear decoding problems, and multi-cell scenarios. The authors of [121] develop a stochastic geometry model for dense MTC systems adopting massive MIMO setups however, their focus is on providing a random access solution for such networks, failing to analyze the impact of massive MIMO during the data transmission phase. Additionally, it is worth noticing that all these works assume that all devices are synchronized to the UL of the base station.

Therefore, we decided to focus our work on the data transmission phase, by proposing a solution where clusters of MTC devices share exclusive and periodic time-frequency resources and simultaneously transmit their data with massive MIMO technology being deployed at the BS to retrieve each one of the device's transmissions. By using massive MIMO at the BS, a great number of MTC devices can be assigned to the same time-frequency resources, consequently, mitigating the negative effects on HTC, *e.g.*, data congestion. The proposed solution allows the addition of MTC services to current wireless cellular networks without the necessity of allocating additional time-frequency resources. The proposed transmission scheme uses only a small fraction of the time-frequency resources, and therefore, it has a small impact on the ability of the BS to serve HTC devices.

3.3 The UL Mixed-service Communications Problem

In this section, we investigate the feasibility of massive MIMO as a means to address the so-called UL mixed-service communication problem. We propose an approach that enables a huge number of bursty and low rate devices in a cell without compromising

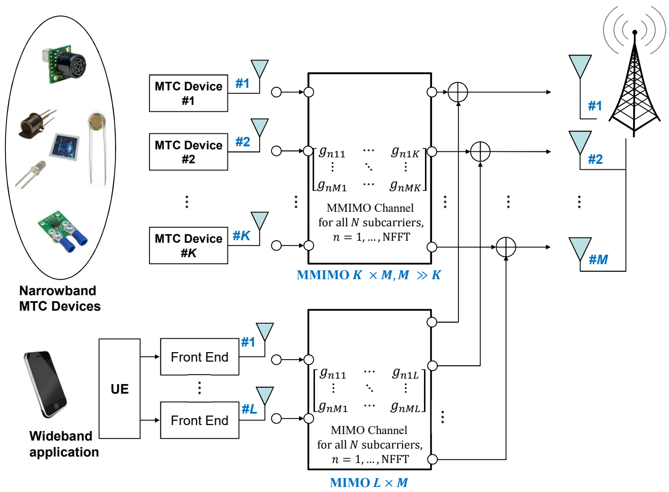

Figure 3.1 – Context application: Enabling a great number of low rate MTC devices in a cell.

the Radio Access Network (RAN) as depicted in Figure 3.1. Our proposal is in line with the set of MTC features considered in 3rd Generation Partnership Project (3GPP) [118, 122]:

(i) low mobility: the devices rarely move or only move within a certain region;
(ii) time-controlled: MTC data delivery only occurs during predefined time intervals;
(iii) time tolerant: MTC data transfer can be delayed;
(iv) small data transmissions: only small amounts of data are exchanged between the device and the BS, *i.e.*, bursty transmissions;
(v) mobile-originated only: MTC devices utilizing only mobile originated communications;
(vi) infrequent transmission: long period between two data transmissions.

Treating MTC devices as regular UEs turns out to be an issue, as scheduling PRBs in extremely dense networks is a nontrivial task made harder in the presence of retransmissions and intrinsic UL synchronization procedures [50–53].

Therefore, in order to ease the scheduling task and increase the network capacity as well, in this work, we propose the creation of a new channel, named PNSCH, which has its time-frequency resources shared with a great number of MTC devices. The main idea behind the PNSCH is the exploitation of the multi-path propagation characteristic offered by the wireless communications channels and harness it by employing the massive MIMO technology [4]. Therefore, a BS equipped with a multitude of antennas can use the spatial degrees of freedom offered by the wireless channel to separate the data stream of MTC devices.

This way, the individual data streams conveyed by spatially spread MTC signals can be separated thanks to the inherent spatial multiplexing properties of massive MIMO technology [4], where the antenna array size at the BS is at least one order of

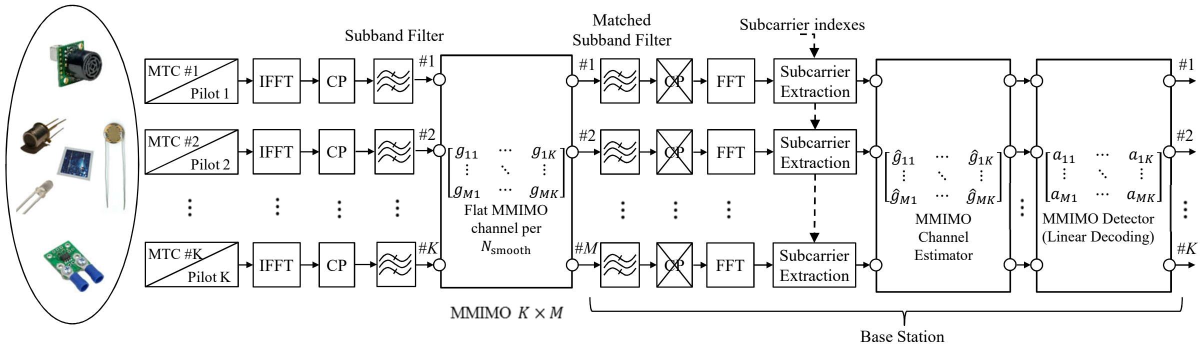

Figure 3.2 – Block diagram of a Massive MIMO UL for mixed networks, where the BS simultaneously serves narrowband MTC devices and wideband UEs (*i.e.*, HTC devices). The cluster of MTC devices seen at the transmit side share the same time-frequency PRBs, while the sole BS at the receive side is equipped with an antenna array at least one order of magnitude larger than the number of MTC devices.

magnitude larger than the number of served MTC devices. Next, we describe the system depicted in Figure 3.2 in terms of its underlying functional blocks.

3.3.1 Signal Generation & Transmission

We assume the transmitted signals of a cluster with K single-antenna MTC devices are detected by a Massive MIMO BS equipped with M receive antennas, $M \gg K$. All the K MTC sources map data into a set of continuous PRBs in the frequency domain, with the subcarrier indexes providing the spectral position of the PNSCH at the physical layer level.

As the focus of our work is on the data transmission phase (*i.e.*, during the RRC_CONNECTED state [49]), we, therefore, assume that all MTC devices being served by a BS are already synchronized and connected to it before accessing the PNSCH, *i.e.*, the MTC devices have already performed the random access and attach procedures before any data is sent through the PNSCH. Before any transmission, in order to align its UL transmissions to the BS timing, each one of the MTC devices must perform a random access procedure through the PRACH [49, 52, 98]. Upon successful random access procedure, an MTC device holds a Cell-Radio Temporary Identifier (C-RNTI) that is then mapped to a pilot sequence, which will be used uniquely by that device while it is connected to the BS. The MTC device will use the same pilot sequence whenever it needs to transmit data towards the BS. This unique correspondence between an MTC device and a pilot sequence guarantees orthogonality among all the MTC devices being served by the same BS, which is of utmost importance to massive MIMO systems due to the pilot-contamination problem that might arise when pilot sequences are reused [79]. The interested reader is referred to [48, 94, 95] for a list of solutions to the random access prob-

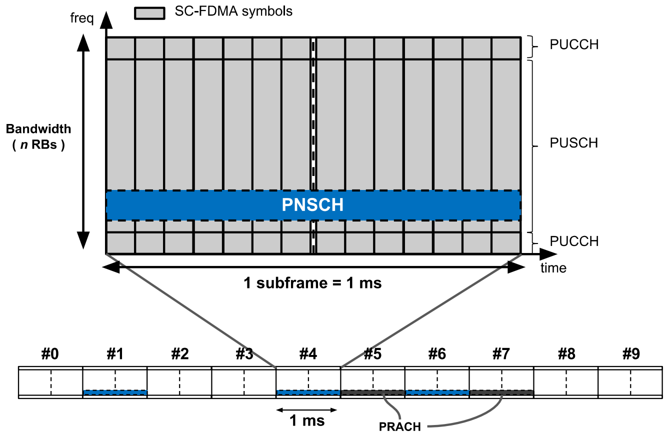

Figure 3.3 – UL Frame Structure with PNSCH.

lem posed by a large number of random access attempts coming from a massive number of MTC devices.

The BS broadcasts SIBs, just like it is done for the PRACH used in current 4G systems (see, *e.g.* [98] and the references therein), in order to configure the PNSCH at the MTC devices. This allows the number of PNSCH transmission opportunities in the UL to be scheduled while taking into consideration discrepancies between the (likely different) capacities of MTC devices and UEs. PNSCH time-frequency resources are semi-statically allocated by the BS and repeat periodically. Additionally, the SIB messages can carry, for instance, information about the pilot sequence length, which in turn, dictates the capacity of the PNSCH as it will determine the remaining time destined to data symbols. The pilot sequence length can be varied so that more MTC devices can be simultaneously served by the BS at the cost of smaller data capacity. Figure 3.3 depicts the UL frame structure devised for the PNSCH. As can be seen in the figure, we assume 1 ms long PNSCH transmission opportunities. The PNSCH is time- and frequency-multiplexed with PUSCH, PUCCH and PRACH as illustrated in the figure.

In this chapter, we assume that inter-cell interference is negligible. Inter-cell interference can be heavily mitigated, and therefore, considered insignificant, if less-aggressive frequency-reuse (*e.g.*, reuse of 3 or 7) is adopted [101]. Inter-cell interference manifests itself in two ways, namely, coherent and non-coherent interference, being the former caused by contaminating cells (*i.e.*, cells that use the same set of pilots as the home cell, causing pilot-contamination) and the latter caused by non-contaminating cells (*i.e.*, cells that do not use the same pilots as the home cell) [41]. In multi-cell scenarios, pilot-contamination, and consequently, coherent interference, can be disregarded once the PNSCH time-frequency resource intervals in each one of the neighbor cells can be configured to refrain them from overlapping with the intervals chosen for the target cell. This kind of configuration can be implemented in order to improve the overall system per-

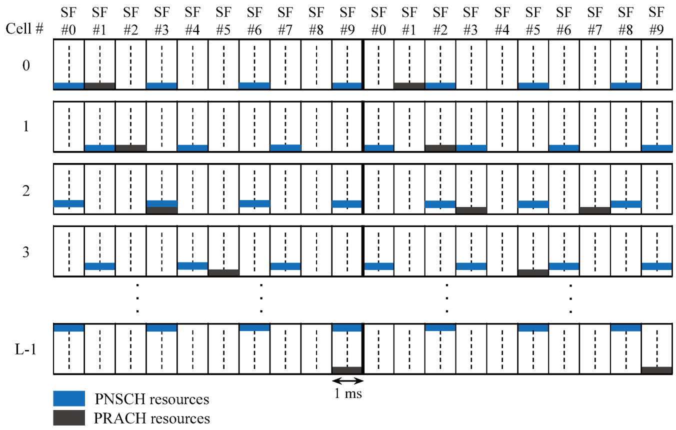

Figure 3.4 – Possible PNSCH resource configuration throughout different neighbor cells.

formance as pilot-contamination results in degradation of the channel estimate quality, which directly impacts on the SE [41]. Figure 3.4 illustrates one possible configuration for the PNSCH intervals of neighbor cells so that pilot-contamination is mitigated.

We assume the utilization of OFDM block-based transmissions. We denote the OFDM symbol interval by T_s, the subcarrier spacing by Δf, the useful symbol duration by $T_u = 1/\Delta f$, and the guard interval (duration of the cyclic prefix) by $T_g = T_s - T_u$. As defined in (2.1), we call the reciprocal of the guard interval, when measured in subcarrier spacings, the *frequency smoothness interval*, N_{smooth}. As showed in the previous chapter, within the same coherence block, a total of τ_p OFDM symbols are used for pilot sequence transmission and τ_u symbols are used for data transmission.

3.3.1.1 Pilot Transmission

As widely used in LTE systems [49], we adopt Zadoff-Chu sequences to design the mutually orthogonal pilot sequences allocated to the MTC devices. These sequences present unit-norm elements but also the additional feature that each sequence is the cyclic shift of another sequence [123, 124]. However, any other set of sequences exhibiting the mutual orthogonality property could be used as pilot sequence, *e.g.*, Walsh Hadamard sequences [125]. Within a cell, each terminal is assigned a $\tau_p N_{\text{smooth}}$ pilot sequence, which is orthogonal to the pilot sequences that are assigned to other terminals in the cell. Collectively, the $K \leq \tau_p N_{\text{smooth}}$ terminals in the cell have the pilot book represented by Φ - a $\tau_p N_{\text{smooth}} \times K$ unitary matrix such that $\Phi^H \Phi = I_K$. The pilot sequence assigned to the k-th MTC device is represented by the column vector ϕ_k.

3.3.1.2 Data transmission

We assume that the modulated symbols (carrying data of an MTC device) are randomly and independently drawn from a digital modulation alphabet (*e.g.*, PSK,

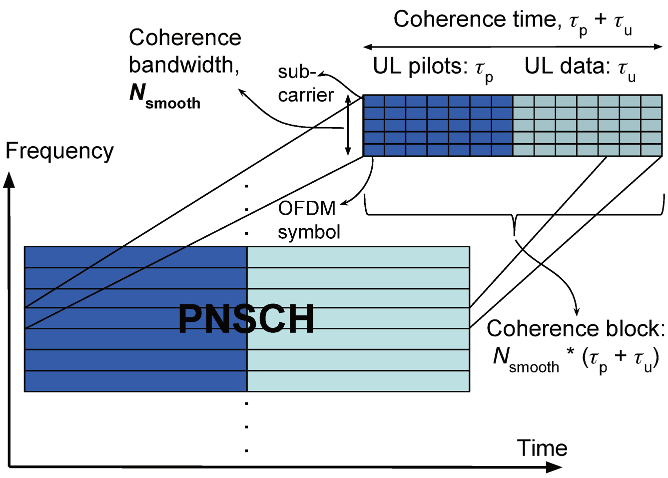

Figure 3.5 – Time-frequency plane.

16-QAM, etc.) with normalized average energy. The modulated symbols are mapped into τ_u OFDM symbols.

Each MTC device transmits its signal (*i.e.*, allocated pilot sequence and data) by taking the IFFT of the mapped pilot sequence and data, and subsequently adding a CP.

Figure 3.5 shows the time-frequency plane for one possible configuration of the PNSCH. It shows how pilots and data symbols are mapped into the frequency-/time-domain of a coherence block. In that configuration, half of the OFDM symbols are used for pilots and the other half for data transmission. The time-frequency plane is divided into coherence blocks in which each channel is time-invariant and frequency-flat. The fraction of pilot symbols and UL data can be selected based on the network traffic characteristics, *i.e.*, the PNSCH configuration can be modified to increase the number of served MTC devices or the data rate per MTC device by increasing the number of OFDM symbols used for pilot or data transmission accordingly.

As already mentioned in chapter 2, it is seen in Figure 3.2, that filters are added to the transmission and reception chains in order to mitigate OOB emissions and consequently, provide a better coexistence between PNSCH and other UL channels.

3.3.2 The Massive MIMO Channel

In this chapter, we consider the same channel model thoroughly described in 2.3.2. As presented in that chapter, the $M \times K$ channel matrix G can be compactly expressed as $\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}$.

3.3.3 Linear MMSE Channel Estimation

Here we consider the case where CSI, *i.e.*, $\mathbf{G}$ is estimated from the received pilot sequences at the BS. As mentioned earlier, we do not consider the existence of pilot-contamination during the channel estimation phase once the PNSCH intervals in all cells (target and neighbor ones) can be configured to avoid transmission overlapping.

3.3.3.1 De-Spreading of the Received Pilot Signal

The pilot sequences propagate through the UL channel and are received by the BS as a $M \times \tau_p N_{\text{smooth}}$ signal,

$$\mathbf{Y} = \sqrt{\rho}\ \mathbf{G}\ \Phi^H + \mathbf{W}, \tag{3.1}$$

where ρ is the average UL transmit power of each MTC device (UL Tx power) and the elements of the $M \times \tau_p N_{\text{smooth}}$ receiver noise matrix, $\mathbf{W}$, are i.i.d. $\mathcal{CN}(0, 1)$. As the noise variance is assumed to be equal to 1, thus, ρ can be interpreted as a normalized transmit SNR and consequently is dimensionless. The BS performs a de-spreading operation of the received signal by correlating it with each one of the K pilot sequences. This is the equivalent of right-multiplying the received signal matrix by the k-th pilot sequence, ϕ_k, resulting in

$$\begin{aligned} \mathbf{y}_k &= \frac{\mathbf{Y}\ \phi_k}{\sqrt{\rho}} \\ &= \mathbf{G}\ \Phi^H\ \phi_k + \frac{\mathbf{W}\ \phi_k}{\sqrt{\rho}} \\ &= \mathbf{g}_k + \mathbf{w'}, \end{aligned} \tag{3.2}$$

where $\mathbf{w'} = \frac{\mathbf{W}\ \Phi_k}{\sqrt{\rho}}$ is an $M \times 1$ noise vector, whose elements are i.i.d. $\mathcal{CN}(0, 1/\rho)$ because they are related to the original Gaussian noise matrix by a unitary multiplication scaled by $1/\sqrt{\rho}$. The de-spread signal, $\mathbf{y}_k$, is Gaussian distributed as follows

$$\mathbf{y}_k \sim \mathcal{CN}\left(\mathbf{0}_M, \left[\beta_k + \frac{1}{\rho}\right] \mathbf{I}_M\right). \tag{3.3}$$

Remark 2. *As $\rho \to \infty$, the variance of $\boldsymbol{y}_k \to \beta_k$, i.e., $\boldsymbol{y}_k \to \boldsymbol{g}_k$.*

Equation (3.2) is also known as the LS estimator [55],

$$\hat{\mathbf{g}}_k^{\text{LS}} = \mathbf{y}_k, \tag{3.4}$$

and its mean-square estimation error per antenna is denoted by

$$\eta_k^{\text{LS}} = \frac{1}{M}\mathbb{E}[\|\tilde{\mathbf{g}}_k^{\text{LS}}\|^2] = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_k^{\text{LS}} - \mathbf{g}_k\|^2] = \frac{1}{\rho}. \tag{3.5}$$

The LS channel estimation error is correlated with both the channel estimate and the de-spread received signal,

$$\frac{1}{M}\mathbb{E}[(\tilde{\mathbf{g}}_k^{\text{LS}})^H \hat{\mathbf{g}}_k^{\text{LS}}] = \frac{1}{M}\mathbb{E}[(\tilde{\mathbf{g}}_k^{\text{LS}})^H \mathbf{y}_k] = \frac{1}{\rho}. \tag{3.6}$$

As the LS channel estimate, $\hat{\mathbf{g}}_k^{\text{LS}}$, the estimation error $\tilde{\mathbf{g}}_k^{\text{LS}}$ is also Gaussian distributed as follows,

$$\tilde{\mathbf{g}}_k^{\text{LS}} \sim \mathcal{CN}\left(\mathbf{0}_M, \frac{1}{\rho}\mathbf{I}_M\right). \tag{3.7}$$

Next, we present the MMSE estimator, which exhibits better performance than that of the LS estimator [79].

3.3.3.2 Channel Estimation

After de-spreading, the BS has a noisy version of the channel vector, which is defined by (3.2). Under the assumption of independent Rayleigh fading, the elements of the channel vector and the noise vector are statistically independent. By assumption, the large-scale fading coefficients are considered known at the BS, so the prior distribution of $g_{m,k} \sim \mathcal{CN}(0, \beta_k)$, is also known. In section 3.4 we outline a possible approach for estimation of the large-scale fading coefficients at the BS. The linear MMSE estimator is the estimator achieving minimum MSE among all estimators of such form [44]. That is, it solves the following the optimization problem

$$\hat{\mathbf{g}}_k^{\text{MMSE}} = \arg \min_{\mathbf{B} \in B^{M \times M}} \|\mathbf{g}_k - \mathbf{B}\mathbf{y}_k\|^2, \tag{3.8}$$

where $\mathbf{y}_k$ is defined in (3.2), $\mathbf{B}$ is a matrix that minimizes the MSE. After solving (3.8), we find that the linear MMSE channel estimator is given by

$$\begin{aligned} \hat{\mathbf{g}}_k^{\text{MMSE}} &= \frac{\beta_k}{\beta_k + \frac{1}{\rho}}\mathbf{y}_k \\ &= \frac{\rho\beta_k}{\rho\beta_k + 1}\mathbf{y}_k. \end{aligned} \tag{3.9}$$

Note that as $\rho \to \infty$, the MMSE estimator becomes the LS estimator. The mean-square per antenna of the channel estimate is denoted by γ_k and given by

$$\gamma_k = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_k^{\text{MMSE}}\|^2] = \frac{\rho\beta_k^2}{1 + \rho\beta_k}. \tag{3.10}$$

The channel estimation error is denoted by

$$\tilde{\mathbf{g}}_k^{\text{MMSE}} = \hat{\mathbf{g}}_k^{\text{MMSE}} - \mathbf{g}_k, \tag{3.11}$$

and the mean-square estimation error per antenna of the MMSE estimator is

$$\begin{aligned}\eta_k^{\text{MMSE}} = \frac{1}{M}\mathbb{E}[\|\tilde{\mathbf{g}}_k^{\text{MMSE}}\|^2] = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_k^{\text{MMSE}} - \mathbf{g}_k\|^2] \\ = \frac{\beta_k}{1+\rho\beta_k} \\ = \beta_k - \gamma_k.\end{aligned} \tag{3.12}$$

Remark 3. *As $\rho \to \infty$, the MSE per antenna, $\eta_k^{MMSE} \to 0$.*

The channel estimation error is uncorrelated with both the channel estimate and the de-spread received signal,

$$\frac{1}{M}\mathbb{E}[(\tilde{\mathbf{g}}_k^{\text{MMSE}})^H \hat{\mathbf{g}}_k^{\text{MMSE}}] = 0. \tag{3.13}$$

$$\frac{1}{M}\mathbb{E}[(\tilde{\mathbf{g}}_k^{\text{MMSE}})^H \mathbf{y}_k] = 0. \tag{3.14}$$

The estimation error $\tilde{\mathbf{g}}_k^{\text{MMSE}}$ and the estimate $\hat{\mathbf{g}}_k^{\text{MMSE}}$ are jointly Gaussian distributed as follows,

$$\hat{\mathbf{g}}_k^{\text{MMSE}} \sim \mathcal{CN}\left(\mathbf{0}_M, \gamma_k \mathbf{I}_M\right). \tag{3.15}$$

$$\tilde{\mathbf{g}}_k^{\text{MMSE}} \sim \mathcal{CN}\left(\mathbf{0}_M, (\beta_k - \gamma_k)\,\mathbf{I}_M\right). \tag{3.16}$$

Therefore, the fact that they are uncorrelated implies that they are also statistically independent.

Remark 4. *As $\rho \to \infty$, the variance of $\hat{\boldsymbol{g}}_k^{MMSE} \to \beta_k$.*

One final remark about the MMSE estimation is that the MMSE estimator of a Gaussian random variable, $\mathbf{g}_k$, that is observed in independent Gaussian noise, **w'**, is a linear estimator and thus equals the linear MMSE channel estimator, *i.e.*, there exists no better non-linear Bayesian estimator in this special case [44].

3.3.4 Linear Detection

In this section, we present the studied linear detectors when imperfect channel knowledge is considered. As shown in chapter 2, linear detectors perform fairly well and asymptotically achieves capacity when $M \to \infty$. [4]. Therefore, we consider again linear detection methods in our subsequent study.

3.3.4.1 MRC

Given imperfect channel knowledge and MRC detection, the detection matrix, $\mathbf{A}$, is defined as

$$\mathbf{A}_{\text{MRC}} = \hat{\mathbf{G}} \tag{3.17}$$

The received signal associated with the k-th MTC device when using the MRC detector is defined as

$$r_k = \hat{\mathbf{g}}_k^H \mathbf{y} = \sqrt{\rho}\|\hat{\mathbf{g}}_k\|^2 x_k + \sqrt{\rho} \sum_{i=0, i\neq k}^{K} \hat{\mathbf{g}}_k^H \mathbf{g}_i x_i + \hat{\mathbf{g}}_k^H \mathbf{w}, \tag{3.18}$$

where $\hat{\mathbf{g}}_k$ is the k-th column of the matrix $\mathbf{A}_{\text{MRC}} = \hat{\mathbf{G}}$.

3.3.4.2 ZF

Given imperfect channel knowledge and ZF detection, the linear detection matrix, $\mathbf{A}$, is defined as

$$\mathbf{A}_{\text{ZF}} = \hat{\mathbf{G}}(\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1}. \tag{3.19}$$

3.3.4.3 MMSE

Given imperfect channel knowledge and MMSE detection, the linear detection matrix, $\mathbf{A}$, is defined as

$$\begin{aligned} \mathbf{A}_{\text{MMSE}} &= \hat{\mathbf{G}} \left[\hat{\mathbf{G}}^H \hat{\mathbf{G}} + \frac{1}{\rho} \text{cov}\left(\mathbf{w} - \sqrt{\rho} \tilde{\mathbf{G}} \mathbf{x} \right) \right]^{-1} \\ &= \hat{\mathbf{G}} \left[\hat{\mathbf{G}}^H \hat{\mathbf{G}} + \frac{1}{\rho} \left(1 + \rho \sum_{l=1}^{K} \mathbb{E}\left[\tilde{\mathbf{g}}_l \tilde{\mathbf{g}}_l^H \right] \right) \mathbf{I}_K \right]^{-1}, \end{aligned} \tag{3.20}$$

where $\text{cov}(a)$ denotes the covariance matrix of a random variable a and $\tilde{\mathbf{G}}$ is the estimation error, $\mathbf{G} = \hat{\mathbf{G}} - \tilde{\mathbf{G}}$.

3.3.5 Achievable Rates

In this subsection, we derive lower bounds on the achievable rate for each one of the studied linear detectors when MMSE channel estimation is considered. Considering that during one PNSCH transmission interval at the target cell there are no other PNSCH transmissions being originated at other cells or that a less-aggressive frequency-reuse factor (*e.g.*, reuse factor of 3 or 7) is adopted, we can analyze and derive the achievable rates as if the target cell was a single-cell system, emphasizing the fact that inter-cell interference is inexistent or negligible and therefore, do not need to be accounted for. The rationale behind the single-cell MU-MIMO analysis is that its results are readily comprehended, they bound the performance of multi-cell systems and that the single-cell performance can be actually achieved if less-aggressive frequency-reuse is adopted. For the following derivations, we use standard capacity bounding techniques from the massive MIMO literature [41, 101, 115].

The received signal vector at the BS is defined as

$$\begin{aligned}\mathbf{r} = \mathbf{A}^H\mathbf{y} = \mathbf{A}^H\left(\sqrt{\rho}\mathbf{G}\mathbf{x} + \mathbf{w}\right) &= \mathbf{A}^H\left(\sqrt{\rho}\hat{\mathbf{G}}\mathbf{x} - \sqrt{\rho}\tilde{\mathbf{G}}\mathbf{x} + \mathbf{w}\right) \\ &= \sqrt{\rho}\mathbf{A}^H\hat{\mathbf{G}}\mathbf{x} - \sqrt{\rho}\mathbf{A}^H\tilde{\mathbf{G}}\mathbf{x} + \mathbf{A}^H\mathbf{w},\end{aligned} \tag{3.21}$$

where $\tilde{\mathbf{G}} = \hat{\mathbf{G}} - \mathbf{G}$ is the estimation error matrix. Due to the properties of the MMSE estimation, $\tilde{\mathbf{G}}$ is independent of $\hat{\mathbf{G}}$, see (3.13). Therefore, the received signal associated with the k-th MTC device is given by

$$\begin{aligned} r_k &= \sqrt{\rho}\mathbf{a}_k^H\hat{\mathbf{G}}\mathbf{x} - \sqrt{\rho}\mathbf{a}_k^H\tilde{\mathbf{G}}\mathbf{x} + \mathbf{a}_k^H\mathbf{w} \\ &= \underbrace{\sqrt{\rho}\mathbf{a}_k^H\hat{\mathbf{g}}_k x_k}_{\text{desired signal}} + \underbrace{\sqrt{\rho}\sum_{l=1,l\neq k}^{K}\mathbf{a}_k^H\hat{\mathbf{g}}_l x_l}_{\text{intra-cell interference}} \underbrace{- \sqrt{\rho}\sum_{l=1}^{K}\mathbf{a}_k^H\tilde{\mathbf{g}}_l x_l + \mathbf{a}_k^H\mathbf{w}}_{\text{effective noise}},\end{aligned} \tag{3.22}$$

where $\mathbf{a}_k$, $\hat{\mathbf{g}}_k$ and $\tilde{\mathbf{g}}_k$ are the k-th columns of $\mathbf{A}$, $\hat{\mathbf{G}}$, and $\tilde{\mathbf{G}}$, respectively. As $\tilde{\mathbf{G}}$ and $\hat{\mathbf{G}}$ are independent, $\mathbf{A}$ and $\tilde{\mathbf{G}}$ are also independent. As shown in (3.22), the BS treats the channel estimate as the true channel, and the last three terms in the equation are considered as intra-cell interference and effective noise respectively. Therefore, the achievable UL rate for the k-th MTC device is defined by (2.20), where the SINR is given by

$$\text{SINR}_k = \frac{\rho|\mathbf{a}_k^H\hat{\mathbf{g}}_k|^2}{\rho\sum_{l=1,l\neq k}^{K}|\mathbf{a}_k^H\hat{\mathbf{g}}_l|^2 + \rho\|\mathbf{a}_k\|^2\sum_{l=1}^{K}\eta_l^{\text{MMSE}} + \|\mathbf{a}_k\|^2}, \tag{3.23}$$

where η_l^{MMSE} is defined in (3.12). By following a similar line of reasoning as in [41, 101] we obtain lower bounds on the achievable rate for the MRC, ZF, and MMSE linear detectors.

3.3.5.1 MRC detector

The received SINR of the k-th stream for the MRC detector is defined as

$$\begin{aligned}\text{SINR}_k^{\text{MRC}} &= \frac{\rho\|\hat{\mathbf{g}}_k\|^4}{\rho\sum_{l=1,l\neq k}^{K}|\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_l|^2 + \rho\|\hat{\mathbf{g}}_k\|^2\sum_{l=1}^{K}\left[\mathbb{E}\left[\tilde{\mathbf{g}}_l\tilde{\mathbf{g}}_l^H\right]\right]_{ll} + \|\hat{\mathbf{g}}_k\|^2} \\ &\to \frac{\|\mathbf{g}_k\|^4}{\sum_{l=1,l\neq k}^{K}|\mathbf{g}_k^H\mathbf{g}_l|^2}, \text{ as } \rho \to \infty.\end{aligned} \tag{3.24}$$

The second line of (3.24) follows directly from Remarks 2 and 3, showing that $\hat{\mathbf{g}}_l \to \mathbf{g}_l$, $\forall\, l$ as $\rho \to \infty$.

Considering MMSE channel estimation, i.i.d. Rayleigh fading, and MRC processing, the achievable rate for the k-th MTC device is lower bounded by

$$\tilde{R}_k^{\mathrm{MRC}} = \log_2\left(1 + \frac{\rho(M-1)\gamma_k}{1 + \rho\sum_{l=1}^{K}\beta_l - \rho\gamma_k}\right). \tag{3.25}$$

If M grows without limit and we consider that the transmit power of each MTC user can be scaled with M according to $\rho = P/\sqrt{M}$, where P is fixed, then (3.25) becomes

$$\begin{aligned}\tilde{R}_k^{\mathrm{MRC}} &= \log_2\left(1 + \frac{\frac{\frac{P^2\beta_k^2(M-1)}{M}}{1+\frac{P}{\sqrt{M}}\beta_k}}{1 + \frac{P}{\sqrt{M}}\sum_{l=1}^{K}\beta_l - \frac{\frac{P^2\beta_k^2}{M}}{1+\frac{Pd_k}{\sqrt{M}}}}\right) \\ &\to \log_2\left(1 + P^2\beta_k^2\right), M \to \infty.\end{aligned} \tag{3.26}$$

This result shows that with a large number of antennas and imperfect CSI, which is obtained through MMSE channel estimation based on UL pilots, available at the BS, the intra-cell interference vanishes and the transmit power of each MTC device can only be reduced proportionally to $1/\sqrt{M}$. This is due to the fact that when the transmitted power of each MTC device is cut down, both data and pilot signals suffer from a reduction in power. A *squaring effect* happens at the receiver since data and pilot signals are multiplied together, and therefore, as a consequence of this *squaring effect*, it is not possible to reduce the transmitted power proportionally to $1/M$ as in the perfect channel knowledge case.

The result in (3.26) is equal to the performance of a interference-free (*i.e.*, intra-cell interference) SISO link with transmit power $P^2\beta_k^2$ and with no small-scale fading (*i.e.*, fast-fading). Additionally, the use of a large number of antennas at the BS increases the SE of the system K times by simultaneously serving K MTC devices over the same time-frequency resources. As we show in the sequence, the same reasoning is true for ZF and MMSE detectors.

Remark 5. *After analyzing* (3.25) *and* (3.26)*, we see that if we decrease the transmit power proportionally to* $1/M^{\alpha}$*, where* $\alpha > 1/2$*, then the SINR of the UL transmission from the k-th device will go to zero as* $M \to \infty$*. When* $\alpha < 1/2$ *the SINR of the UL transmission from the k-th device grows without bound as* $M \to \infty$*. This means that* $1/\sqrt{M}$ *(i.e.,* $\alpha = 1/2$*) is the fastest rate at which we can decrease the transmit power of each device and still maintain a fixed rate.*

3.3.5.2 ZF detector

The received SINR of the k-th stream for the ZF detector is defined as

$$\begin{aligned}\mathrm{SINR}_k^{\mathrm{ZF}} &= \frac{\rho}{\left(1+\rho\sum_{l=1}^{K}\left[\mathbb{E}\left[\tilde{\mathbf{g}}_l\tilde{\mathbf{g}}_l^H\right]\right]_{ll}\right)\left[\left(\hat{\mathbf{G}}^H\hat{\mathbf{G}}\right)^{-1}\right]_{kk}} \\ &\to \frac{\rho}{\left[\left(\mathbf{G}^H\mathbf{G}\right)^{-1}\right]_{kk}}, \text{ as } \rho\to\infty.\end{aligned} \tag{3.27}$$

The second line of (3.27) follows directly from Remarks 2 and 3.

Considering MMSE channel estimation, i.i.d. Rayleigh fading, ZF processing, and that $M \geq K+1$, the achievable rate for the k-th MTC device is lower bounded by

$$\tilde{R}_k^{\mathrm{ZF}} = \log_2\left(1+\frac{\rho\gamma_k(M-K)}{1+\rho\sum_{l=1}^{K}\eta_l^{\mathrm{MMSE}}}\right). \tag{3.28}$$

If M grows without limit and we make $\rho = P/\sqrt{M}$, then (3.28) becomes

$$\begin{aligned}\tilde{R}_k^{\mathrm{ZF}} &= \log_2\left(1+\frac{\frac{\frac{P^2\beta_k^2(M-K)}{M}}{1+\frac{Pd_k}{\sqrt{M}}}}{1+\frac{P}{\sqrt{M}}\sum_{l=1}^{K}\left(\beta_l-\frac{\frac{Pd_l^2}{\sqrt{M}}}{1+\frac{Pd_l}{\sqrt{M}}}\right)}\right) \\ &\to \log_2\left(1+P^2\beta_k^2\right), M\to\infty.\end{aligned} \tag{3.29}$$

3.3.5.3 MMSE detector

The received SINR of the k-th stream for the MMSE receiver is defined as

$$\begin{aligned}\mathrm{SINR}_k^{\mathrm{MMSE}} &= \frac{1}{\left[\left(\mathbf{I}_K+\frac{\rho}{1+\sum_{l=1}^{K}\left[\mathbb{E}\left[\tilde{\mathbf{g}}_l\tilde{\mathbf{g}}_l^H\right]\right]_{ll}}\hat{\mathbf{G}}^H\hat{\mathbf{G}}\right)^{-1}\right]_{kk}}-1 \\ &\to \frac{\rho}{\left[\left(\mathbf{G}^H\mathbf{G}\right)^{-1}\right]_{kk}}, \text{ as } \rho\to\infty.\end{aligned} \tag{3.30}$$

The second line of (3.30) follows directly from Remarks 2 and 3, and shows that as the SNR increases, ZF and MMSE detectors will present the same SE (see Figure 2.8).

Considering MMSE channel estimation, i.i.d. Rayleigh fading, MMSE processing, the achievable rate for the k-th MTC device is approximately lower bounded by

$$\tilde{R}_k^{\mathrm{MMSE}} = \log_2\left(1+(\alpha_k-1)\theta_k\right), \tag{3.31}$$

where

$$\alpha_k = \frac{(M-K+1+(K-1)\mu)^2}{M-K+1+(K-1)\kappa}, \tag{3.32}$$

$$\theta_k = \frac{M-K+1+(K-1)\kappa}{M-K+1+(K-1)\mu} w\gamma_k, \tag{3.33}$$

where $w = \left[\frac{1}{\rho} + \sum_{l=1}^{K} \eta_l^{\text{MMSE}}\right]^{-1}$, μ and κ are obtained by solving the following two equations:

$$\mu = \frac{1}{K-1} \sum_{l=1, l\neq k}^{K} \frac{1}{Mw\gamma_l \left(1 - \frac{K-1}{M} + \frac{K-1}{M}\mu\right) + 1} \tag{3.34}$$

$$\begin{aligned} &\kappa \left(1 + \sum_{l=1, l\neq k}^{K} \frac{w\gamma_l}{\left(Mw\gamma_l \left(1 - \frac{K-1}{M} + \frac{K-1}{M}\mu\right) + 1\right)^2}\right) \\ &\quad = \sum_{l=1, l\neq k}^{K} \frac{w\gamma_l \mu + 1}{\left(Mw\gamma_l \left(1 - \frac{K-1}{M} + \frac{K-1}{M}\mu\right) + 1\right)^2} \end{aligned} \tag{3.35}$$

3.4 Estimation of Large-Scale Fading Coefficients

In this section, we describe how the large-scale fading coefficients, $\beta_k, \forall k$, can be estimated based on the de-spread signal, $\mathbf{y}_k$. We employ the ML method to estimate next-generation large-scale fading coefficients [126]. Applying the ML method to $f(\mathbf{y}_k; \beta_k)$ with distribution defined in (3.3), we find the following estimator for β_k given the observation $\mathbf{y}_k$

$$\hat{\beta}_k = \frac{\|\mathbf{y}_k\|^2}{M} - \frac{1}{\rho}. \tag{3.36}$$

This estimator exhibits a central Chi-square distribution with $2M$ degrees of freedom. It has $\mathbb{E}[\hat{\beta}_k] = \beta_k$, which shows that the ML estimator is unbiased, and $\text{var}(\hat{\beta}_k) = \frac{(\beta_k + \frac{1}{\rho})^2}{M}$. The mean-square error of the proposed large-scale fading coefficient estimator is defined as

$$\mathbb{E}[(\beta_k - \hat{\beta}_k)^2] = \frac{\left(\beta_k + \frac{1}{\rho}\right)^2}{M}. \tag{3.37}$$

Note that the mean-square error of the proposed estimator is also equal to its variance.

Remark 6. *As* $M \to \infty$, $\mathbb{E}[(\beta_k - \hat{\beta}_k)^2] = var(\hat{\beta}_k) \to 0$.

Remark 6 shows that as M increases, the estimator defined in (3.36) becomes a deterministic value and that it tends, in the limit, to the actual β_k value once the mean-square error vanishes.

In order to assess the efficiency of the estimator we derive the Cramér-Rao bound as [126]

$$\text{var}(\hat{\beta}_k) \geq \frac{(\beta_k + \frac{1}{\rho})^2}{M}. \tag{3.38}$$

Therefore, as can be noticed, the ML estimator derived for β_k is the Minimum Variance Unbiased Estimator (MVUE), *i.e.*, it is an unbiased estimator that has the lowest variance among all other possible unbiased estimators [126].

Additionally, we show that the proposed estimator approaches β_k as the number of antennas M increases.

$$\lim_{M\to\infty} \hat{\beta}_k = \lim_{M\to\infty} \frac{\|\mathbf{y}_k\|^2}{M} - \frac{1}{\rho} \overset{\text{a.s.}}{=} \beta_k. \tag{3.39}$$

The proof of (3.39) is provided in Appendix B. This is an example of the strong law of large numbers and can be interpreted as the variations of $\frac{\|\mathbf{y}_k\|^2}{M}$ becoming increasingly concentrated around its mean value $\mathbb{E}\left[\frac{\|\mathbf{y}_k\|^2}{M}\right] = \beta_k + \frac{1}{\rho}$ as more antennas are added.

3.5 Proposed Channel Estimator

In this section, we propose a simple and practical channel estimator based on the estimator for the large-scale fading coefficients defined in Section 3.4. Our proposed approach estimates the large-scale fading coefficients, β_k, and replaces it into the MMSE channel estimator defined in (3.9), resulting in the following channel estimator

$$\hat{\mathbf{g}}_k^{\text{prop}} = \left(1 - \frac{M}{\rho}\frac{1}{\|\mathbf{y}_k\|^2}\right)\mathbf{y}_k. \tag{3.40}$$

Remark 7. *The proposed estimator asymptotically approaches the ideal MMSE channel estimator in* (3.9) *as* $M \to \infty$,

$$\lim_{M\to\infty} \left(1 - \frac{1}{\rho}\frac{M}{\|\boldsymbol{y}_k\|^2}\right)\boldsymbol{y}_k = \left(\frac{\rho\beta_k}{\rho\beta_k + 1}\right)\boldsymbol{y}_k. \tag{3.41}$$

This can be easily proved by applying Lemma 4 presented in Appendix B to (3.40). The proposed estimator has $\mathbb{E}\left[\hat{\mathbf{g}}_k^{\text{prop}}\right] = \mathbf{0}_M$ and covariance matrix given by

$$\begin{aligned} \text{Cov}[\hat{\mathbf{g}}_k^{\text{prop}}] &= \mathbb{E}[\hat{\mathbf{g}}_k^{\text{prop}}(\hat{\mathbf{g}}_k^{\text{prop}})^H] \\ &= \frac{(\rho\beta_k+1)(\rho\beta_k-1)(M-1)+M}{\rho(\rho\beta_k+1)(M-1)}\mathbf{I}_M \\ &= \left[\frac{\rho\beta_k^2}{(\rho\beta_k+1)} + \frac{1}{\rho(\rho\beta_k+1)(M-1)}\right]\mathbf{I}_M \\ &= \left(\gamma_k+\epsilon_k\right)\mathbf{I}_M, \end{aligned} \tag{3.42}$$

where $\epsilon_k = \frac{1}{\rho(\rho\beta_k+1)(M-1)}$, which is the error introduced by the estimation of β_k.

Remark 8. *As ρ and $M \to \infty$, the variance of $\hat{\boldsymbol{g}}_k^{prop} \to \beta_k$.*

As can be seen by analyzing (3.42), as $M \to \infty$, $\text{Cov}[\hat{\mathbf{g}}_k^{\text{prop}}] \to \frac{\rho\beta_k^2}{1+\rho\beta_k}\mathbf{I}_M$, which is the covariance matrix of the MMSE estimator. The mean-square estimation error per antenna of the proposed channel estimator is defined as

$$\begin{aligned} \eta_k^{\text{prop}} &= \frac{1}{M}\mathbb{E}[\|\tilde{\mathbf{g}}_k^{\text{prop}}\|^2] = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_k^{\text{prop}} - \mathbf{g}_k\|^2] \\ &= \frac{1}{\rho}\left(\frac{M}{M-1}\frac{1}{1+\rho\beta_k} - 1 + 2\theta_k\right), \end{aligned} \tag{3.43}$$

where $\theta_k = \int_0^1\int_{-1}^1 \frac{\kappa_k^2(1-t)+\kappa_k w\sqrt{t(1-t)}}{\kappa_k^2(1-t)+2\kappa_k w\sqrt{t(1-t)}+t} f_W(w) f_T(t) dw dt$ and $\kappa_k \triangleq \sqrt{\rho\beta_k}$ and $f_W(w)$ and $f_T(t)$ are defined by

$$f_T(t) = \frac{\Gamma(2M)}{(\Gamma(M))^2}\left[t(1-t)\right]^{M-1}, \quad 0 < t < 1, \tag{3.44}$$

$$f_W(w) = \frac{M}{\pi}B\left(\frac{1}{2}, M\right)\left(1-w^2\right)^{M-\frac{1}{2}}, \quad |w| < 1. \tag{3.45}$$

The proof of the mean-square estimation error per antenna is given in Appendix C. The analytical mean-square estimation error expression (3.43) is useful for system design and performance evaluation purposes [127].

Remark 9. *The mean-square error between $\hat{\boldsymbol{g}}_k^{prop}$ and $\hat{\boldsymbol{g}}_k^{MMSE}$ is defined as*

$$\frac{1}{M}\mathbb{E}[\|\hat{\boldsymbol{g}}_k^{prop} - \hat{\boldsymbol{g}}_k^{MMSE}\|^2] = \frac{1}{\rho(\rho\beta_k+1)(M-1)} = \epsilon_k. \tag{3.46}$$

The proof of (3.46) is given in Appendix D. From (3.46) we observe that the mean-square error decreases when M, ρ and β_k increase, which consequently shows that the proposed channel estimator asymptotically approaches the performance of MMSE channel estimator.

Next, we also present a simpler and more tractable expression for the mean-square estimation error per antenna, which is defined as

$$\begin{aligned}\frac{1}{M}\mathbb{E}[\|\tilde{\mathbf{g}}_k^{\text{prop}}\|^2] = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_k^{\text{prop}} - \mathbf{g}_k\|^2] \approx \eta_k^{\text{prop(approx.)}} &= \\ = \frac{1}{\rho}\left[\frac{\rho\beta_k}{1+\rho\beta_k} + \frac{1}{(M-1)(1+\rho\beta_k)}\right] & \\ = \beta_k - \gamma_k + \epsilon_k. &\end{aligned} \tag{3.47}$$

Remark 10. *As M or $\rho \rightarrow \infty$, $\eta_k^{prop(approx.)} = \beta_k/(1+\rho\beta_k) = \beta_k - \gamma_k$.*

Remark 10 clearly shows that the approximated mean-square estimation error per antenna of the proposed estimator tends to that of the MMSE estimator when $M \rightarrow \infty$. The proof for the approximated mean-square estimation error per antenna of the proposed channel estimator is given in Appendix E.

The channel estimation error is correlated with the channel estimate and uncorrelated with the de-spread received signal,

$$\frac{1}{M}\mathbb{E}[(\tilde{\mathbf{g}}_k^{\text{prop}})^H \hat{\mathbf{g}}_k^{\text{prop}}] = \epsilon_k. \tag{3.48}$$

$$\frac{1}{M}\mathbb{E}[(\tilde{\mathbf{g}}_k^{\text{prop}})^H \mathbf{y}_k] = 0. \tag{3.49}$$

Remark 11. *As M or $\rho \rightarrow \infty$, then $\frac{1}{M}\mathbb{E}[(\tilde{\boldsymbol{g}}_k^{prop})^H \hat{\boldsymbol{g}}_k^{prop}] = 0$.*

The estimation error, $\tilde{\mathbf{g}}_k^{\text{prop}}$, has the following mean vector and covariance matrix,

$$\mathbb{E}[\tilde{\mathbf{g}}_k^{\text{prop}}] = \mathbf{0}_M, \tag{3.50}$$

$$\text{Cov}\,[\tilde{\mathbf{g}}_k^{\text{prop}}] = (\beta_k - \gamma_k + \epsilon_k)\mathbf{I}_M. \tag{3.51}$$

3.5.1 Complexity Analysis

The computational complexity is a important factor indicating the efficiency of a channel estimator. In this section, we assess the computational complexity of the studied channel estimators. Table 3.1 summarizes the complexities involved in the calculation of the estimators in terms of number of floating-point operations (flops) and the Big-$\mathcal{O}$ notation. The Big-$\mathcal{O}$ notation, also called Landau's symbol, which is a well-known symbolism widely used in complexity theory to describe the asymptotic behavior of algorithms [128]. It basically indicates how fast an algorithm grows or declines. In the table P is the length of the pilot sequence (*i.e.*, $\tau_p N_{\text{smooth}}$) and $\mathbf{D}$ is a $K \times K$ identity matrix

Table 3.1 – Complexities involved in the studied linear channel estimators.

Estimator	Operation	Flops	$\mathcal{O}(.)$
LS	$\mathbf{Y}\Phi$	$MK(2P-1)$ or $2MKP$ if P is large	MKP
MMSE/Prop.	$\mathbf{Y}\Phi\mathbf{D}$	$MK(2P-1)+MK(2K-1)$ or $2MK(P+K)$ if P and K are large	$MK(P+K)$
	Calculation of the elements in $\mathbf{D}$	MMSE: Elements are considered perfectly known	-
		Prop.: calculation of $\lVert\mathbf{y}_k\rVert^2\ \forall k$: $K(2M-1)$ flops or $2KM$ if M is large	MK

with the diagonal elements equal to $\frac{\rho\beta_k}{\rho\beta_k+1}\ \forall k$ and $\left(1-\frac{M}{\rho}\frac{1}{\lVert\mathbf{y}_k\rVert^2}\right)\ \forall k$ for the MMSE and proposed channel estimators respectively. It is important to notice that in this work and in the great majority of works in the literature [69, 79, 129, 130] the large scale fading coefficients, $\{\beta_k\}$, are assumed perfectly known for the MMSE channel estimation.

The LS estimator is the most computational cost-efficient among all of them, presenting a complexity of $\mathcal{O}(MPK)$, however, as will be shown later, this is the least efficient estimator in terms of MSE. The MMSE estimator is the most efficient in terms of MSE, exhibiting a complexity of $\mathcal{O}(MK(P+K))$, however, as mentioned earlier, the complexity involved in the calculation of large-scale fading coefficients (*i.e.*, the elements of $\mathbf{D}$) is not taken into account. On the other hand, the proposed estimator presents MSE efficiency that asymptotically approaches that of the MMSE estimator and has a complexity of $\mathcal{O}(MK(P+K+1))$, where the calculation (*i.e.*, estimation) of the large-scale fading coefficients is already considered in the presented complexity.

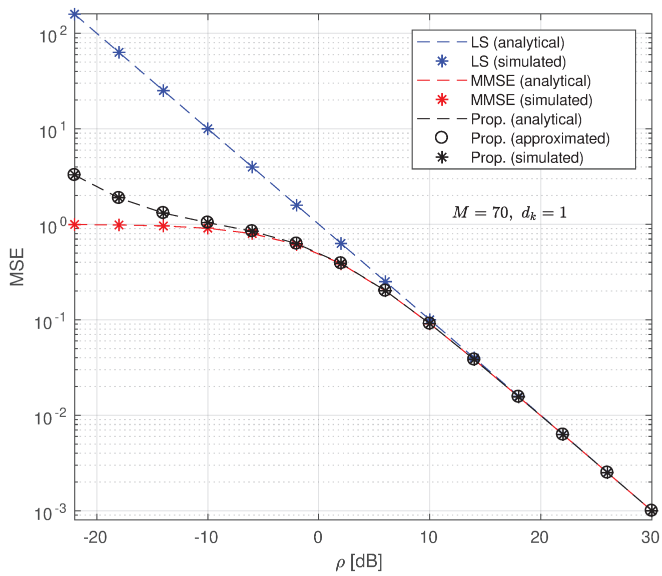

Figure 3.6 – Channel estimation MSE versus average UL pilot power, ρ.

3.6 Simulation Results

In this section, we compare the performance of the proposed channel estimator with that of LS and MMSE channel estimators. Additionally, we assess the performance of MRC, ZF, and MMSE linear decoders when the MMSE and the proposed channel estimators are employed. The performance of each linear decoder is quantified in terms of its BER over a range of UL Tx power (*i.e.*, ρ) values.

We consider two different types of simulation setups for the large-scale fading coefficient, β_k, one with fixed values and other with random values. For the fixed case, we set $\beta_k = 1$. For the random case, the MTC devices in the cell (see Figure 2.4) are uniformly distributed within a ring with radii $r_0 = 100$ m and $r_1 = 1000$ m respectively. The large-scale fading coefficients $\{\beta_k\}$ are independently generated by $\beta_k = \psi / \left(\frac{r_k}{r_0}\right)^v$, where $v = 3.8$, $10 \log_{10}(\psi) \sim \mathcal{N}(0, \sigma^2_{\text{shadow,dB}})$ with $\sigma_{\text{shadow,dB}} = 8$, and r_k is the distance of the k-th MTC device to the BS. Both, the path loss exponent, v, and the standard deviation of the log-normal shadowing, $\sigma_{\text{shadow,dB}}$, are common values for outdoor shadowed urban cellular radio environments [103,116]. For all simulations we assume $K = 10$ MTC devices.

Figure 3.6 shows the MSE versus UL Tx power results (ρ) for the case when $\beta_k = 1$ and $M = 70$. As can be noticed, the analytical, approximated and simulated MSE curves match for all the studied channel estimators. As expected, the MSE of all estimators decreases as ρ increases. At low ρ values (values lower than -10 dB), the MSE of the proposed estimator is higher than that of the MMSE estimator, however, it is still smaller than that of the LS estimator. On the other hand, with the increase of ρ (for values higher than -10 dB), the gap between the MMSE and proposed estimators decreases.

The MSE versus the number of BS antennas, M, is compared in Figure 3.7 for the case when $\beta_k = 1$ and $\rho = 10$ dB. The MSE of the proposed estimator decreases as M increases, approaching the MSE of the MMSE estimator, while the MSE of the LS and MMSE channel estimators stay constant. The result showed in the figure is in

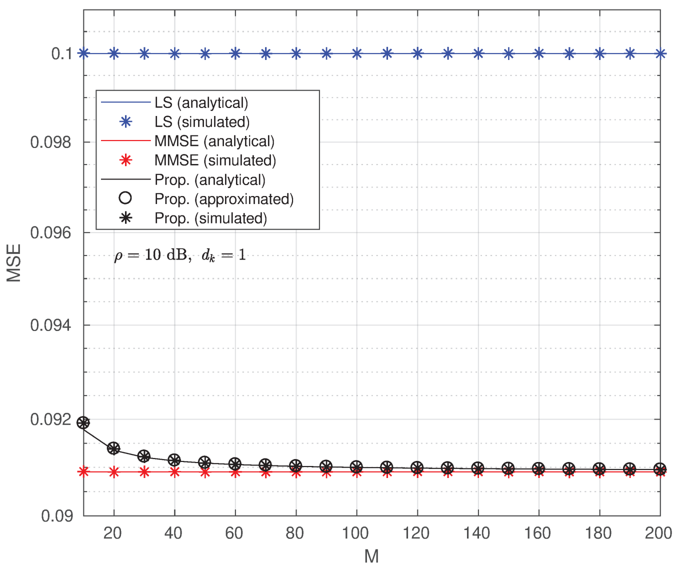

Figure 3.7 – MSE performance versus number antennas, M.

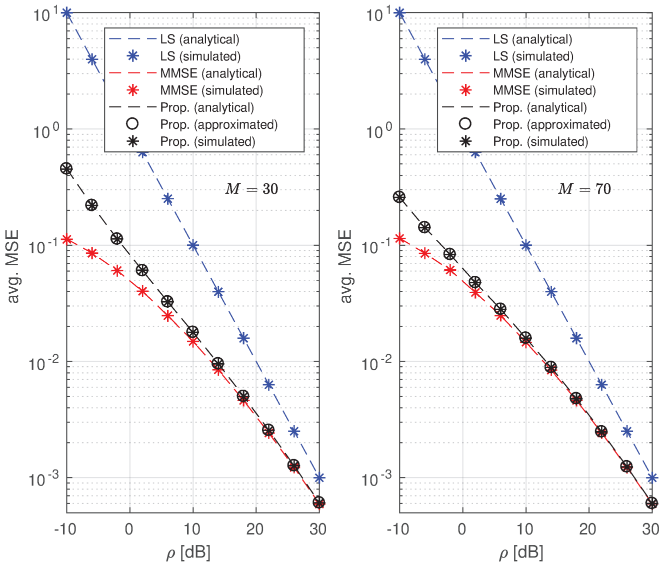

Figure 3.8 – Average channel estimation MSE under random β_k versus UL pilot power, ρ.

line with Remark 7. Additionally, it is also worth mentioning that the approximated MSE expression for the proposed channel estimator matches the values of the closed-form expression.

In Figure 3.8, we assess the variation of the MSE under random large-scale fading $\{\beta_k\}$ for $M = 30$ and $M = 70$ respectively. The results are obtained by averaging the MSE values over 10^4 realizations of $\{\beta_k\}$. As can be noticed, the simulated MSE values match the values of the analytical and approximated MSE expressions. It can be also noticed that the MSE of the proposed channel estimator approaches that of the MMSE channel estimator as M increases. The proposed channel estimator outperforms the LS channel estimator significantly at low UL Tx power values. It is important to notice

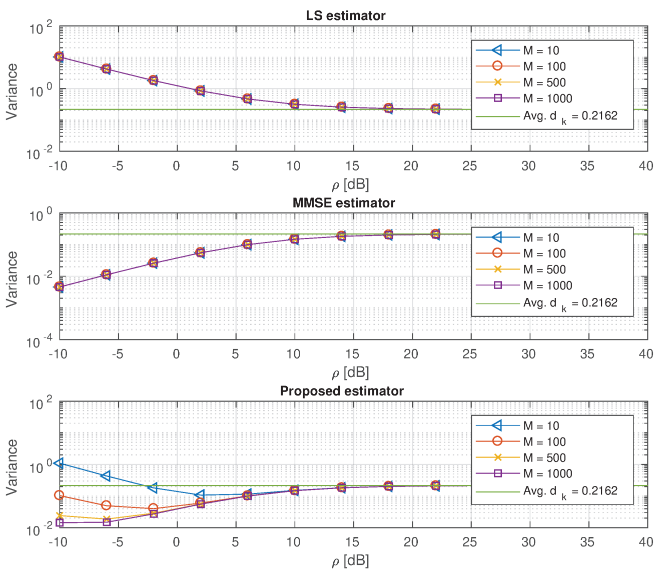

Figure 3.9 – Comparison of the averaged variances of the channel estimators.

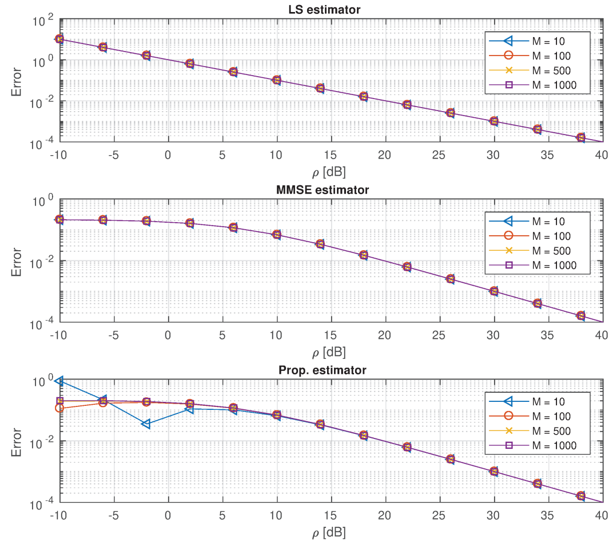

Figure 3.10 – Average of the absolute error between the large-scale fading, β_k, and the variance of the studied estimators.

that the MSE of both LS and MMSE estimators does not depend on M and therefore, does not vary as M varies.

Figures 3.9 and 3.10 present the averaged variance and error under random large-scale fading coefficient, $\{\beta_k\}$, respectively. The results are obtained by averaging the variance and error values over 10^7 different realizations of $\{\beta_k\}$. As can be seen, the variance and consequently the error of the LS and MMSE channel estimators do not depend on M (both of them depend only on ρ), however, the variance of the proposed channel estimator depends on both M and ρ. As M increases, the variance curve for next-generation proposed channel estimator approaches that of the MMSE channel estimator. It is also worth mentioning that the variance of both MMSE and proposed channel estimators

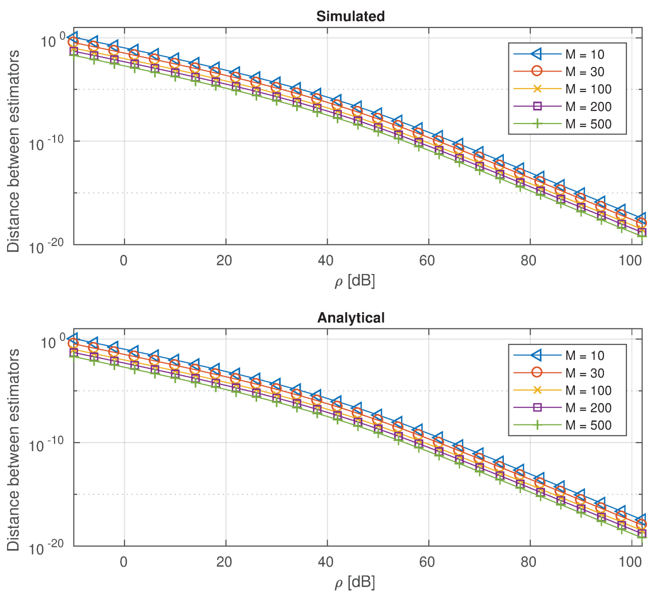

Figure 3.11 – Averaged distance between proposed and MMSE channel estimators.

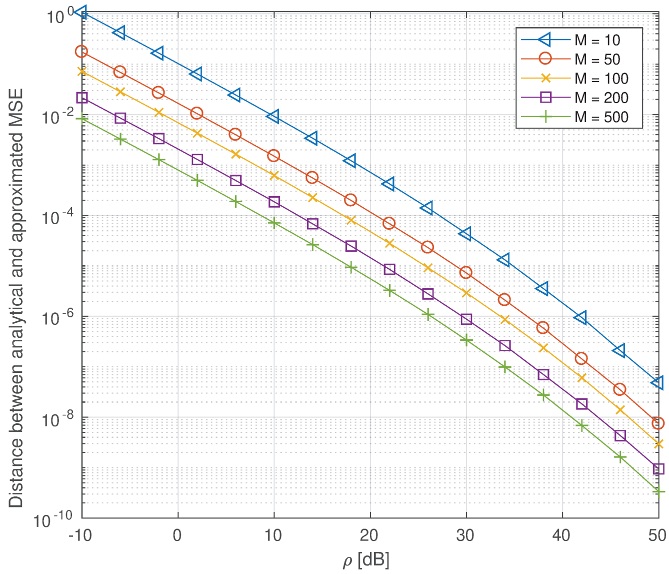

Figure 3.12 – Averaged absolute distance between closed-form and approximated MSE error expressions, $|\eta_k^{\text{prop}} - \eta_k^{\text{prop(approx.)}}|$.

converges faster to the average β_k than the variance of the LS estimator as can be seen in Figure 3.10. These results are in line with Remarks 2, 4 and 8, showing that the variance of all the studied channel estimators tend to the average β_k.

The averaged distance between the proposed and MMSE channel estimators for different number of antennas, M, and UL Tx power values, ρ, under random large-scale fading, $\{\beta_k\}$, is depicted in Figured 3.11. The results are obtained by averaging the MSE values over 10^4 realizations of $\{\beta_k\}$ for the simulated and closed-form distances between the two estimators. As stated in Remark 9, the distance between the estimators decreases as M and ρ increase.

In Figure 3.12, we compare the averaged absolute distance between the approximated MSE expression presented in (3.43) and the analytical (closed-form) MSE expression presented in (3.47) under random large-scale fading $\{\beta_k\}$ for various UL Tx power (*i.e.*, ρ) and M values. The results are obtained by averaging the absolute distance between the MSE expressions over 10×10^3 realizations of $\{\beta_k\}$. The distance between the MSE expressions is large (around 1.07) at low ρ, decreasing with ρ as expected. For $M = 100$ and $\rho = -10$ dB the averaged absolute error between the expressions is around 0.072. The results presented in Figure 3.12 show that the approximated MSE expression expression can be used instead of the more complex one given by (3.43) when M is at least one order of magnitude larger than K (*i.e.*, $M \gg K$) and/or at high UL Tx power regimes.

Figure 3.13 shows lower bounds and simulated SE for MRC, ZF, and MMSE detectors employing MMSE channel estimation under random large-scale fading $\{\beta_k\}$.

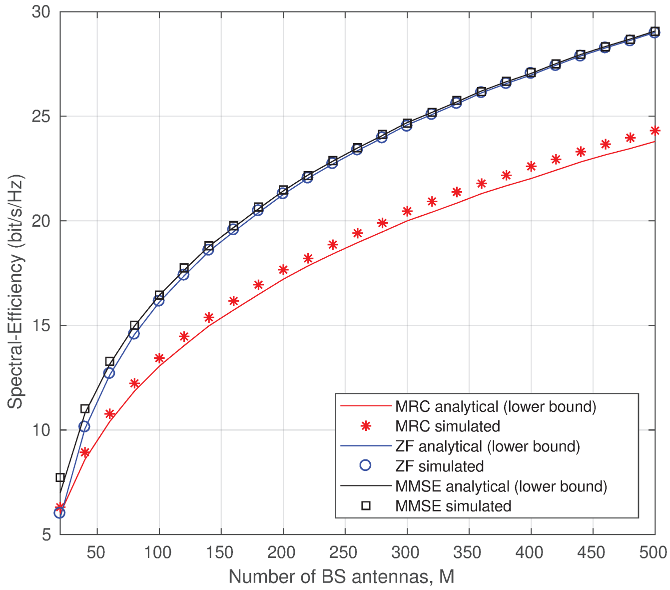

Figure 3.13 – Lower bounds and numerically evaluated values of the SE for different number of BS antennas for MRC, ZF, and MMSE detectors.

The SE is averaged over 10^6 realizations of $\{\beta_k\}$. In this simulation ρ is set to 10 dB and there are $K = 10$ MTC devices. As can be seen, at this UL Tx power, the SE for $M = 500$ is in the order of $24-29$ bits/s/Hz, corresponding to a SE of $2.4-2.9$ bits/s/Hz per MTC device. These values are in line with practical values, for example, 64-QAM with rate 1/2 channel coding corresponds to 3 bits/s/Hz [101]. As can be seen from the figure, the MMSE detector is always better than the MRC and ZF detectors, however, the performance of the ZF detector asymptotically approaches the MMSE with increasing M. The conclusion here is that even with simple and sub-optimal linear processing such as MRC and ZF it is possible to achieve high SE and consequently serve more MTC devices at the same time/frequency resource. It is also worth discussing the achievable rates when the proposed estimator is employed instead of the MMSE one. As it has been shown earlier, the performance of the proposed estimator asymptotically approaches that of the MMSE estimator as M and/or ρ increases, therefore, the lower bounds derived here can be thought of as upper bounds for linear detectors using the proposed estimator, *i.e.*, they will never perform better than linear detectors using the MMSE channel estimator once at best, the proposed estimator will be equal to the MMSE one.

In Figure 3.14 we show the transmit power per user that is needed to reach a fixed SE. The figure shows the normalized power, ρ, required to achieve 1 and 2 bits/s/Hz per user as a function of M. It can be seen that by doubling the number of antennas, M, the transmit power can be cut back by approximately 2 dB. When M is large, the difference in performance between MRC and ZF (or MMSE) is less than 1.5 dB. This difference increases when the target SE is increased. For the 2 bits/s/Hz case, the cross-talk interference, *i.e.*, interference from other devices, is more significant (relative to thermal

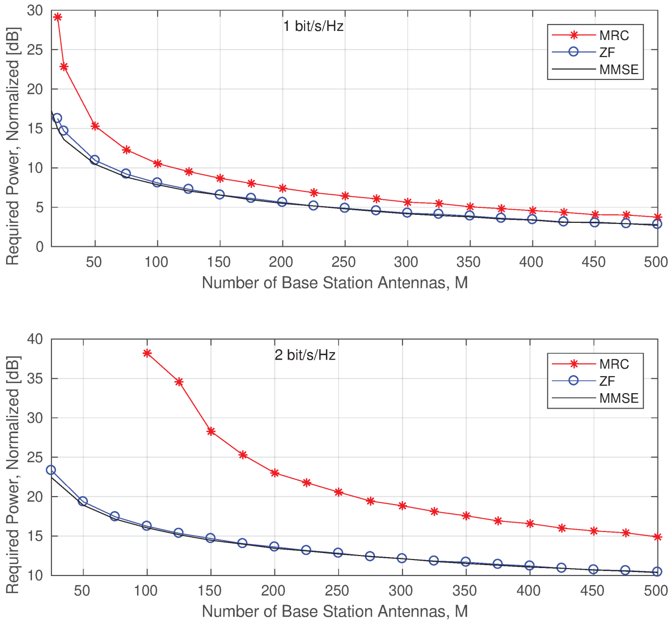

Figure 3.14 – Transmit power required to achieve 1 and 2 bits/s/Hz per user for MRC, ZF, and MMSE linear receivers as a function of the number of antennas M. The number of users is set to $K = 10$, and the propagation parameters are $\sigma_{\text{shadow}} = 8$ dB and $v = 3.8$.

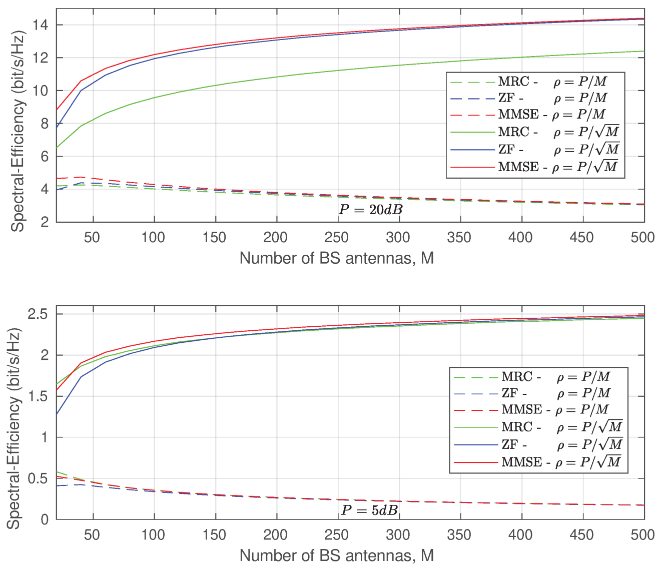

Figure 3.15 – SE versus number of antennas, M, for linear detectors: MRC, ZF, and MMSE. In these examples $K = 10$ MTC devices are served simultaneously and the transmit power is $P = 20$ dB for the upper figure and $P = 5$ dB for the lower figure and the propagation parameters are $\sigma_{\text{sshadow}} = 8$ dB and $v = 3.8$.

noise) and therefore, ZF and MMSE receivers perform relatively better. The difference in performance between MRC and ZF (or MMSE) for $M = 500$ antennas is approximately 4.47 dB. This characteristic of massive MIMO is very important to MTC devices, where stringent power constraints have to be respected.

In Figure 3.15 we illustrate the power-scaling laws. The figure shows the SE versus the number of antennas M for $\rho = P/M$ and $\rho = P/\sqrt{M}$ and with MRC, ZF, and MMSE linear detectors respectively. In the upper part of the figure, we set P to 20 dB. At this SNR, the SE is in the order of 4-14 bits/s/Hz, which corresponds to a SE per MTC device of 0.4-1.4 bits/s/Hz. As expected, with $\rho = P/M$, when the number of antennas, M, grows, the SE tends to 0. However, when $\rho = P/\sqrt{M}$, the SE converges to a non-zero limit as $M \to \infty$. Next, in the lower part of the figure, we set P to 5 dB. This figure provides the same insights as the upper one. The gap between the performance of MRC and that showed by the ZF and MMSE detectors is reduced compared next-generation upper part of the figure. This is due to the fact that the interference caused by cross-talk, *i.e.*, the interference coming from other MTC devices, is smaller in this figure when compared to the thermal noise. These results attest that the transmit power of each MTC device can be scaled down as $P/\sqrt{M}$. These results show that the radiated power by MTC device can be made inversely proportional to the number of BS antennas, M, with no reduction in performance. Additionally, as the radiated power levels are reduced, the cross-talk interference caused by the use of the inferior MRC detector eventually drops below the noise level, making this simple linear detector a practical option, once it performs as well as ZF and MMSE detectors for the large M regime.

Next, we present some simulations to asses how the BER behaves when a large number of antennas and sub-optimal linear detectors are employed at the BS. For the following simulation results, we assume OFDM parameters similar to LTE: (i) a symbol interval of $T_s = 500/7 \approx 71.4$ us, (ii) a subcarrier spacing of $\Delta f = 15$ KHz, (iii) a useful symbol duration $T_u = 1/\Delta f \approx 66.7$ us, and (iv) a cyclic prefix interval (guard interval) $T_g = T_s - T_u \approx 4.76$ us. The frequency smoothness interval is approximately $N_{\text{smooth}} = 14$ subcarriers. We assume a coherence time of 1 ms (equivalent to 14 OFDM symbols), of which 1 symbol is used to send UL pilots and the remaining symbols, *i.e.*, 13 symbols, are used to send UL data. Therefore, in this case, the maximum number of MTC devices, $K = \tau_p N_{\text{smooth}} = 14$. Additionally, we consider uncoded QPSK UL data transmissions.

We use the MFB as benchmark for the BER comparisons. The MFB is also known in the literature as the perfect interference-cancellation bound [55]. As the name suggests, MFB performs as the k-th user of a matched-filter receiver in the absence of other sources of interference [55]. Our motivation for this choice is that for $M \gg K$ both multi-user interference and small-scale fading effects tend to disappear (thanks to the processing gains of Massive MIMO), so the performance of the MIMO $K \times M$ channel, which is assumed to be flat Rayleigh fading inside a coherence block, approaches the MFB as $M \to \infty$. The simulation results discussed in the sequel were averaged over 10^{10} realizations. The simulation type is Monte-Carlo with a bit error counting procedure that compares the transmitted bit vector to the received bit vector.

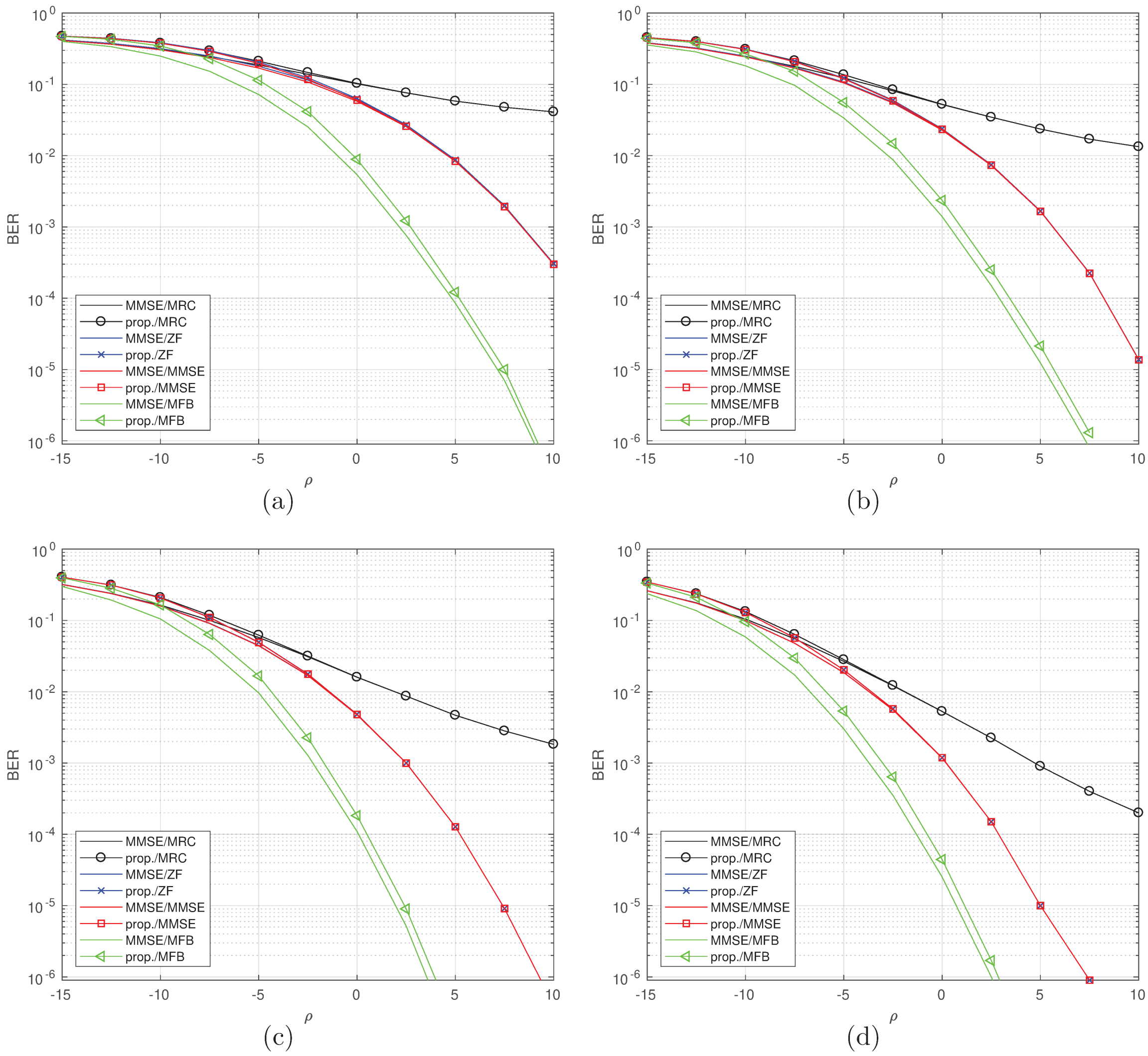

Figure 3.16 – BER performance of different linear detectors methods for $K = 10$ single-antenna MTC devices and different array sizes at the BS. MFB is provided as a benchmark for comparisons. (a) $M = 50$ antennas. (b) $M = 100$ antennas. (c) $M = 250$ antennas. (d) $M = 500$ antennas.

Figure 3.16 shows the BER of the linear filtering detectors described in section 3.3.4 for a fixed number of $K = 10$ MTC devices and BS array sizes in the range of $50 \leq M \leq 500$ antennas. For UL Tx power values smaller than 0 dB, the gap between all the linear detectors (MRC, ZF, and MMSE) employing the proposed channel estimator and the ones employing MMSE estimator is noticeable, vanishing as ρ increases. This gap also decreases as M increases. In the following discussion, we consider the performance of the studied linear detectors without differentiating if they employ MMSE or the proposed channel estimator to calculate the equalization matrix $\mathbf{A}$ once the analysis applies to both cases.

After analyzing the figure, the performance gap inherent to MRC becomes evident, although, as can be seen, it can be dramatically reduced at the expense of larger array sizes at the BS. These results suggest that even low-complexity MRC detectors have the potential to approximate the MFB when M can be made large enough.

As expected, due to its better balance between interference suppression and noise enhancement, the MMSE detector outperforms both the MRC and ZF detectors in all cases studied. The performance gap between ZF and MMSE, which is small enough to be considered negligible for $M \leq 50$, entirely disappears as the BS array size is increased to $M = 250$ or more antennas. As a matter of fact, the main conclusion that can be drawn from the results presented here is that the MRC, ZF, and MMSE detectors all approach the performance of MFB as $M \to \infty$, however, the gap between the MFB and the ZF/MMSE detectors decreases at a faster pace when compared to that of the MRC detector. It is also important to notice that (2.4) is an approximation that only becomes an equality when $M \gg K$, $M \to \infty$, which explains why the MRC detection is not optimum for the results presented here.

3.7 Conclusions

In this chapter, we extended the study presented in Chapter 2, where we propose the use of a Massive MIMO setup as means to tackle the UL mixed-service communication problem, by assuming imperfect channel knowledge. By assuming the availability of a PNSCH, the capacity of the network is increased through the creation of clusters of MTC devices that share the same time-frequency physical resource blocks for data transmission.

We present a simple and practical channel estimator that does not need previous knowledge of the large-scale fading coefficients. The proposed channel estimator presents MSE results that asymptotically approach the ones exhibited by the MMSE channel estimator as the number of antennas and/or UL Tx power increases. Additionally, we derive closed-form and approximated MSE expressions for the proposed estimator, which can be useful in system design and performance evaluation.

Moreover, our simulation results suggest that, as the size of the antenna array at the BS is made progressively larger, the performance of sub-optimal linear detectors approaches the perfect interference-cancellation bound.

4 Channel Estimation for Multi-Cell Massive MIMO Systems Considering Pilot Contamination

4.1 Introduction

Massive MIMO antenna systems potentially allow Base Stations (BSs) to operate with huge improvements in spectral and radiated energy efficiency, using relatively low-complexity linear processing. The higher SE is attained by serving several terminals in the same time-frequency resource through spatial multiplexing, and the increase in energy efficiency is mostly due to the array gain provided by the large set of antennas [42].

The expected massive MIMO improvements assume that accurate channel estimations are available at both the receiver and transmitter for detection and precoding respectively. Additionally, the reuse of frequencies and pilot reference sequences in cellular communication systems causes interference in channel estimation, degrading its performance. Since both the time-frequency resources allocated for pilot transmission and the channel coherence time are limited, the number of possible orthogonal pilot sequences is also limited, and as a consequence, the pilot sequences have to be reused in neighbor cells of cellular systems. Therefore, channel estimates obtained in a given cell get contaminated by the pilots transmitted by the users in other cells [78]. This coherent interference is known in the literature as pilot contamination, *i.e.*, the channel estimate at the base station in one cell becomes contaminated by the pilots of the users from other cells [4]. The contamination not only reduces the quality of the channel estimates, *i.e.*, increases the MSE, but also makes the channel estimates statistically dependent, even though the true channels are statistically independent. Moreover, pilot contamination does not disappear with the addition of more antennas [44]. Compared with conventional MU-MIMO systems, the pilot contamination effect has a more significant impact on massive MIMO systems as it leads to a finite limit in system performance even when the number of BS antennas, M, grows without limit [4, 42, 78].

Massive MIMO systems operating in TDD assume channel reciprocity between UL and DL in order to minimize pilot overhead, transmitting pilot reference signals only in the UL. In this scenario, the pilot overhead cost is proportional to the number of terminals and improved estimation quality can be achieved due to the large number of antennas [43, 69]. Base stations estimate channels usually based on LS [4] or MMSE

[129–131] methods. Besides, inter and intra-cell large-scale fading coefficients are assumed to be perfectly known when applying the MMSE method in the great majority of works [41, 69, 76, 102, 131, 132].

In a real-world network deployment, although changing slowly, the large-scale fading coefficients must be estimated and updated from time to time. Additionally, the estimation error of the large-scale fading coefficients impacts significantly on the performance of UL data decoding and DL transmission (*e.g.*, precoding and beamforming) [133, 134]. Approaches on how to estimate the large-scale fading coefficients are presented in the following pieces of work [102, 133, 135].

The most commonly used analytical massive MIMO channel is the spatially i.i.d. frequency non-selective (flat) fading channel model. Flat fading channels are also known as *amplitude varying channels* and *narrowband channels* as the signal's bandwidth is narrow compared to channel's bandwidth [103]. In this narrowband channel model, the channel gain between any pair of transmit-receive antennas is modeled as a complex Gaussian random variable. This model relies on two assumptions: (i) the antenna elements in the transmitter and receiver being spatially well separated once the more widely spaced (in wavelengths) the antenna elements, the smaller the spatial channel correlation [136, 137], and (ii) the presence of a large number of temporally but narrowly separated multipaths (common in a *rich-scattering* environment), whose combined gain, by the central limit theorem, can be approximated by a Gaussian random variable [137].

Flat fading channels present a channel response that exhibits flat gain and linear phase over a bandwidth (coherence bandwidth) that is greater than the signal's bandwidth. Therefore, all frequency components of the signal will experience the same magnitude of fading, resulting in a scalar channel response. The gain applied to the signal varies over time according to a fading distribution. In this chapter, we consider that the gain applied to the signal passing through this channel will vary randomly, according to a Rayleigh distribution. We additionally assume that the antenna spacing is sufficiently large so that the antennas are uncorrelated. The Rayleigh block-fading model is a simple model that captures the channel variations for frequency-flat channels. In this model, the channel statistics remain constant during a time interval, known as *coherence time*, T_c, and the frequency response is flat within a frequency band, known as *coherence band*, B_c. The statistics of the channel change independently from block to block (*i.e.*, from interval to interval) [4]. We focus our analysis on a single statistically representative coherence interval, *i.e.*, a time-frequency interval/block ($T_c \times B_c$) where the channel statistics are constant [41].

In this chapter, we deal with the channel estimation and pilot contamination problems associated with UL training in Rayleigh block flat-fading channels and understand its impact on the operation of multi-cell Multi User (MU) massive MIMO TDD cellular systems. We propose and evaluate an efficient and practical channel estimator that does not require previous knowledge of inter/intra-cell large-scale fading coefficients (*i.e.*, interference) and noise power. We employ the ML method to find an estimator for the interference plus noise power term in the MMSE channel estimator. We show that this estimator is not only unbiased but also achieves the Crámer-Rao lower bound. We replace this estimator into the MMSE estimator and prove that the performance of the new channel estimator asymptotically approaches that of the MMSE estimator. Simulation results confirm that the performance of the proposed channel estimator approaches that of the ideal MMSE estimator asymptotically with the number M of antennas, *i.e.*, $M \rightarrow \infty$. Additionally, we derive an approximate analytical MSE expression for the proposed channel estimator that is more mathematically tractable and not susceptible to numerical issues.

The remainder of this chapter is divided as follow: First, in section 4.2, some related works are analyzed and discussed, next, in section 4.3, we present the problem structure, signal model adopted for this study and briefly discuss two well-known channel estimators, namely, LS and MMSE linear estimators. Then, in section 4.4, we introduce the proposed channel estimator for flat Rayleigh fading channels. Later, some numerical results are presented in section 4.5 in order to support the effectiveness of the proposed estimator against the well-known linear estimators. Finally, we present our conclusions in section 4.6.

4.2 Related Work

In this section, we survey previous work on channel estimation and pilot contamination mitigation.

A TDD cellular system employing BSs equipped with large numbers of antennas that communicate simultaneously with smaller numbers of cheap, single-antenna terminals through MU-MIMO techniques is proposed in [4]. The author employs LS channel estimation in order to study and evaluate the problems caused by pilot contamination to such systems. He concludes that even when different sets of orthogonal pilots are used in different cells it makes little difference to the resulting Signal-to-Interference Ratio (SIR). This work is the first one to present the massive MIMO concept and identify its intrinsic issues, however, it fails to suggest ways to mitigate the pilot contamination problem.

The impact of pilot contamination on multi-cell systems is studied in [69]. The authors adopt MMSE channel estimation for the analysis of pilot contamination and the achievable rates in a massive MIMO system suffering from such problem. They propose a multi-cell MMSE-based precoding method that mitigates the pilot contamination problem by considering the set of training sequences assigned to the users in the solution of an optimization problem that minimizes the error seen by users in the serving cell and the interference seen by the users in all other cells. Simulation results show that the proposed approach has significant gains over certain single-cell precoding methods such as zero-forcing. In summary, the authors address the pilot contamination problem through a precoding technique and assume that the large-scale fading coefficients are known to all BSs.

MMSE channel estimation is used in [129] to derive approximations of the achievable UL and DL rates with several linear precoders and detectors for realistic system dimensions, *i.e.*, systems where the number of antennas is not extremely large compared to the number of users. Simulation results show that the approximations are asymptotically tight and accurate for realistic systems. The authors do not propose any approach to mitigate the pilot contamination problem, however, they study and evaluate its impact on the achievable rates.

The impact of pilot contamination effect on the achievable UL ergodic rate when using linear detection in multi-cell MU massive MIMO systems under a more realistic physical channel model is assessed in [130]. The authors assume that the channel vectors for different users are correlated, or not asymptotically orthogonal due to the antennas not being sufficiently well separated and/or the propagation environment not offering rich enough scattering. Moreover, they assume that the BS performs MMSE channel estimation based on training sequences received on the UL and a priori knowledge of the large-scale fading coefficients.

In [131], the Polynomial Expansion (PE) technique is applied to channel estimation of massive MIMO systems in order to approximate the MMSE estimator and thereby obtain a new set of low-complexity channel estimators. Conventional MMSE estimators present cubic complexity due to an inversion operation while the estimator proposed in [131] reduces this to square complexity by approximating the inverse by a L-degree matrix polynomial. The proposed estimator achieves near-optimal MSE with low polynomial degrees. However, the statistical knowledge of channel and disturbance parameters at the receiver is assumed in this chapter.

Outer multi-cellular precoding is employed in [102] to devise a method used to eliminate pilot contamination in massive MIMO systems. Each BS performs two levels of precoding, firstly it estimates and shares only the large-scale fading coefficients with a central entity (network controller) which computes the precoding matrices and sends them back to the BSs, i.e, outer precoding. Next, each BS performs local precoding using estimates of the fast-fading vectors, *i.e.*, inner precoding. The proposed approach is shown to completely mitigate the pilot contamination problem, making it possible to construct interference and noise-free multi-cell massive MIMO systems with frequency reuse one and infinite DL and UL Signal-to-Interference-plus-Noise Ratios (SINRs). The proposed method employs MMSE channel estimation, however, the effectiveness of this method lies in the estimation accuracy of the shared large-scale fading coefficients from each BS. The authors also propose a method to estimate the large-scale fading coefficients. As this approach needs to share the large-scale coefficients with the network controller for outer precoding computation, it presents a higher computational complexity than non-cooperative approaches.

The authors in [76], adopt a massive MIMO system model that is based on spatially correlated channels. They devise a covariance aided channel estimation method that exploits the covariance information of both desired and interfering user channels. The Bayesian method is used to derive two different channel estimators (it is also shown that the Bayesian estimators coincide with the MMSE estimators), one for all channels from users in all cells to the target cell and the other one for the channels from users within the target cell. Results show that in the ideal case, where the desired and the interference covariance matrices span distinct subspaces, the pilot contamination effect tends to vanish in the large antenna array case. As a consequence, users with mutually non-overlapping Angle of Arrival (AoA) hardly contaminate each other. Based on their results, the authors propose a coordinated pilot assignment strategy that assigns carefully selected groups of users to identical pilot sequences.

A semi-blind iterative Space-Alternating Generalized Expectation (SAGE) based channel estimation algorithm for massive MIMO systems with pilot contamination is proposed in [132]. The proposed method does not assume a priori knowledge on the large-scale fading coefficients of the interfering cells, employing an estimate obtained from the received signal. The method updates the pilot based MMSE channel estimates iteratively with the help of the SAGE algorithm, which improves the initial estimate with the help of pilot symbols and soft information of the transmitted data. However, as it refines the channel estimates over some iterations starting from an initial MMSE channel estimation, it presents a computational complexity that is higher than the one presented by pure blind and linear estimators.

After surveying the literature on channel estimation and pilot contamination mentioned above, it is clear that, for clarity, in the great majority of studies the authors always assume complete knowledge on large-scale fading coefficients, *i.e.*, path-loss and shadow fading, of the interfering cells, which is not the case in practical deployments of MU Massive MIMO systems. Furthermore, several studies propose solutions that present additional computational complexity in order to mitigate the pilot contamination problem.

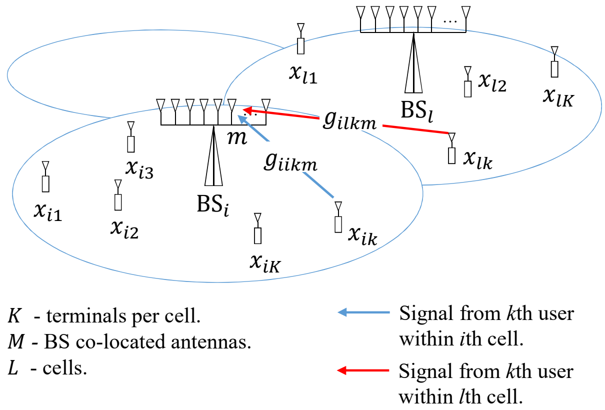

Figure 4.1 – Problem definition.

The main contribution of this chapter is the proposal and assessment of a simple and practical channel estimator used to mitigate the pilot contamination problem. The proposed estimator does not assume a priori knowledge of the large-scale fading coefficients of the interfering cells. Moreover, it does not require the heavy overhead created by their estimation once it obtains them from the received signal.

4.3 Problem Structure

Let's assume, as illustrated in Fig. 4.1, a multi-cell system with L cells, where each cell has a BS at its center with M co-located antenna elements and K randomly located single-antenna users. We assume K users in each cell as it constitutes the worst case for the interference caused by the re-use of pilots in all cells. Let's also assume Rayleigh fading channels being independent across users and antennas. Let g_{ilkm} represent the complex gain of the channel from the k-th user in the l-th cell to the m-th BS antenna in the i-th cell. We can write $g_{ilkm} = \sqrt{\beta_{ilk}} h_{ilkm}$ where $\sqrt{\beta_{ilk}}$ is the large-scale coefficient encompassing both path loss and log-normal shadowing. We assume that h_{ilkm} is the small-scale coefficient with a circularly-symmetric complex normal distribution $\mathcal{CN}(0, 1)$ and that all BS co-located antennas see the same large-scale coefficient value, β_{ilk}. The assumption that the large-scale fading coefficients do not depend on the antenna index m of a given BS as well as on the frequency is due to the fact that typically, the distance between a user and a BS is significantly larger than the distance between the BS antennas [102]. Therefore, between a BS with M co-located antennas and a user, there is only one large-scale fading coefficient, β_{ilk}. Moreover, these coefficients only change when a user considerably change its geographical location. The wireless channels are considered static during the channel coherence time (*i.e.*, channel estimates are effective only in this time interval) and independent across users and antennas.

The $M \times 1$ channel vector from the k-th user in the l-th cell to the M antennas

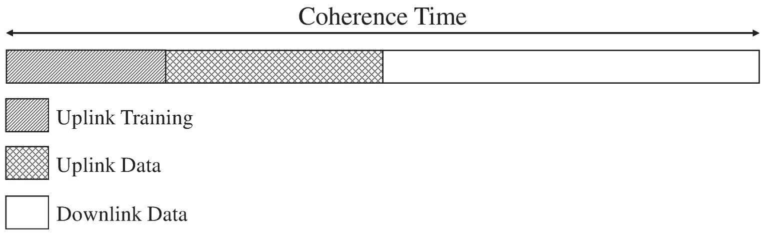

Figure 4.2 – TDD transmission protocol.

at the i-th BS is defined by $\mathbf{g}_{ilk} = [g_{ilk1}, g_{ilk2}, \cdots, g_{ilkM}]^T \sim \mathcal{CN}(\mathbf{0}_M, \beta_{ilk}\mathbf{I}_M)$. The overall $M \times K$ channel matrix $\mathbf{G}_{il}$ is obtained by column concatenating vectors $\mathbf{g}_{ilk}$ for all cell users, that is, $\mathbf{G}_{il} = [\mathbf{g}_{il1}, \mathbf{g}_{il2} \cdots \mathbf{g}_{ilK}]$. For detection and precoding, BS i needs to know the channels of the users in cell i, namely $\{\mathbf{g}_{iik}, \forall k\}$. The same way as in the literature, we treat $\{\beta_{ilk}\}$ as being deterministic during the channel estimation phase [42, 43, 130, 132]. It is possible because the large-scale fading coefficients change slowly in comparison with the small-scale fading coefficients [41]. As described, the overall channel matrix $\mathbf{G}_{il}$ can also be defined directly by the channel coefficients,

$$\begin{aligned}\mathbf{G}_{il} &= \begin{pmatrix} g_{il11} & g_{il21} & \cdots & g_{ilK1} \\ g_{il12} & g_{il22} & \cdots & g_{ilK2} \\ \vdots & \vdots & \ddots & \vdots \\ g_{il1M} & g_{il2M} & \cdots & g_{ilKM} \end{pmatrix} = \mathbf{H}\mathbf{D}^{1/2} \\ &= \begin{pmatrix} h_{il11} & h_{il21} & \cdots & h_{ilK1} \\ h_{il12} & h_{il22} & \cdots & h_{ilK2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{il1M} & h_{il2M} & \cdots & h_{ilKM} \end{pmatrix} \begin{pmatrix} \sqrt{\beta_{il1}} & 0 & \cdots & 0 \\ 0 & \sqrt{\beta_{il2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\beta_{ilK}} \end{pmatrix},\end{aligned} \tag{4.1}$$

where $\mathbf{D}$ is the diagonal matrix with the large-scale coefficients for all K devices, $\beta_{ilk}, k = 1, \ldots, K$.

Based on the assumption of channel reciprocity, we adopt the TDD protocol depicted in Fig. 4.2 and proposed in [138]. Due to the reciprocity principle, only the UL channels need to be estimated while the DL channels are equal to the conjugate-transpose of the UL channels. It is important to note that the length of the TDD frames is limited by the channel coherence time [138, 139]. According to the TDD protocol, first, all users in all cells send their UL training sequences synchronously. After that, the BSs use the training sequences to estimate the UL channels. Next, the users send UL data signals. Then, the BSs use the estimated channels to detect UL data and generate precoding matrices used to transmit DL data.

4.3.1 UL Training

Each user transmits an UL training sequence so that the user serving BS can estimate the channels per antenna and subsequently detect the transmitted user data. We assume that users in different cells transmit data at the same time-frequency resource (a typical scenario in massive MIMO) and that the pilot reuse factor is one, the worst possible use case scenario [4]. As all BSs reuse the same set of pilots and transmit at the same time-frequency resource, the pilot contamination problem arises, consequently, all the other BSs will also receive the pilots sent by users being served by other BSs, limiting the quality of the channel estimation [140].

The pilot signals of K users are represented by a $N \times K$ matrix $\mathbf{\Phi}$ of the form $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_K]$, where N is the length of the pilot sequences. Each pilot sequence is of the form $\boldsymbol{\phi}_k = \left[\boldsymbol{\phi}_k^0, \boldsymbol{\phi}_k^1, \dots, \boldsymbol{\phi}_k^{N-1}\right]^T$. The pilot matrix, Φ, exhibits orthogonal property $\Phi^H \Phi = N\mathbf{I}_K$.

The pilots are created by applying cyclic shifts to ZC root sequences with length N, where N is a prime number. These sequences exhibit some useful properties: (i) cyclically shifted versions of themselves are orthogonal to each other; (ii) constant amplitude; (iii) zero auto-correlation; (iv) flat frequency domain response; (v) cross-correlation between two ZC sequences is low [124]. Some of the reasons why they are adopted in communications systems like LTE are: (i) channel estimation at receiver is made simpler due to their small variation in frequency; (ii) inter-cell interference is reduced as they present low cross-correlation; (iii) high PAPR is reduced due to their small variation in time. ZC sequences are used in this work due to the properties mentioned above [124]; however, any other sequences could be used as long as they exhibit the required orthogonal property. Additionally, we assume that $N \geq K$ in order to avoid underdetermined systems.

The received UL training sequences at the i-th BS can be represented as a $M \times N$ matrix defined as

$$\mathbf{Y}_i = \sqrt{q} \sum_{l=1}^{L} \mathbf{G}_{il} \Phi^H + \mathbf{N}_i, \tag{4.2}$$

where q is the UL power or Transmit-SNR (Tx-SNR) and N_i is a $M \times N$ noise matrix with independent and identically distributed elements following $\mathcal{CN}(0, 1)$.

Equation (4.2) can also be written as showed below, which clearly highlights the coherent inter-cell interference caused by users employing the same pilot sequences in other BSs.

$$\mathbf{Y}_i = \underbrace{\sqrt{q}\mathbf{G}_{ii}\Phi^H}_{\text{Desired pilot signals}} + \underbrace{\sqrt{q} \sum_{l=1, l\neq i}^{L} \mathbf{G}_{il} \Phi^H}_{\text{Undesired pilot signals}} + \underbrace{\mathbf{N}_i}_{\text{Noise}}. \tag{4.3}$$

4.3.2 LS Channel Estimator

For estimation of the channel $\mathbf{g}_{ilk}$ at BS i, a sufficient statistic [126, 141, 142] is given by

$$\begin{aligned} \mathbf{z}_{ik} = \frac{1}{\sqrt{qN}}\mathbf{Y}_i\boldsymbol{\phi}_k &= \sum_{l=1}^{L}\mathbf{g}_{ilk} + \frac{\mathbf{N}_i\boldsymbol{\phi}_k}{\sqrt{qN}} \\ &= \underbrace{\mathbf{g}_{iik}}_{\text{Desired channel}} + \underbrace{\sum_{l=1,l\neq i}^{L}\mathbf{g}_{ilk}}_{\text{Inter-cell interference}} + \underbrace{\frac{\mathbf{N}_i\boldsymbol{\phi}_k}{\sqrt{qN}}}_{\text{Noise}}. \end{aligned} \tag{4.4}$$

where $\mathbf{z}_{ik}$ is a column vector with a $\mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$ distribution and

$$\zeta_{ik} = \sum_{l=1}^{L}\beta_{ilk} + \frac{1}{qN}. \tag{4.5}$$

Additionally, the term corresponding to noise in (4.4) has a $\mathcal{CN}(\mathbf{0}_M, \frac{1}{qN}\mathbf{I}_M)$ distribution.

Therefore, the Least Square estimator is given by [126]

$$\hat{\mathbf{g}}_{iik}^{\text{LS}} = \mathbf{z}_{ik}. \tag{4.6}$$

The estimation error vector, $\tilde{\mathbf{g}}_{iik}^{\text{LS}} = \mathbf{g}_{iik} - \hat{\mathbf{g}}_{iik}^{\text{LS}}$, is distributed as $\tilde{\mathbf{g}}_{iik}^{\text{LS}} \sim \mathcal{CN}\left(\mathbf{0}_M, (\zeta_{ik} - \beta_{iik})\mathbf{I}_M\right)$ and is not uncorrelated with (and consequently not independent of) $\hat{\mathbf{g}}_{iik}^{\text{LS}}$, having a covariance matrix defined as $\text{cov}\left(\hat{\mathbf{g}}_{iik}^{\text{LS}}, \tilde{\mathbf{g}}_{iik}^{\text{LS}}\right) = (\zeta_{ik} - \beta_{iik})\,\mathbf{I}_M$. The MSE per antenna of the LS estimator is given by

$$\eta_{ik}^{\text{LS}} = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_{iik}^{\text{LS}} - \mathbf{g}_{iik}\|^2] = \zeta_{ik} - \beta_{iik}. \tag{4.7}$$

As known, the LS estimator has larger MSE than the MMSE estimator, however, it does not need prior knowledge of the large-scale fading coefficients, $\{\beta_{ilk}\}$.

Remark 12. *Due to pilot contamination, as $q \to \infty$, $\eta_{ik}^{LS} \to \sum_{l=1,l\neq i}^{L}\beta_{ilk}$.*

Remark 13. *If $\boldsymbol{g}_{iik}$ is an i.i.d. complex Gaussian vector, then, we see that $\hat{\boldsymbol{g}}_{ilk}^{LS} = \hat{\boldsymbol{g}}_{iik}^{LS}, \forall l$, meaning that the channel estimates are parallel vectors and therefore, showing the BS's inability to separate users transmitting the same pilot sequence within other cells.*

4.3.3 MMSE Channel Estimator

A great number of massive MIMO works adopt the MMSE estimation method to obtain channel knowledge [69, 130]. Those works assume that all large-scale fading coefficients, *i.e.*, $\{\beta_{ilk}, i \geq 1, l \leq L, 1 \leq k \leq K\}$, are perfectly known. In practice, this assumption might not be reasonable. In case we consider the coefficients $\{\beta_{ilk}\}$ perfectly known at the BS, the ideal MMSE estimator is given by [126]

$$\hat{\mathbf{g}}_{iik}^{\text{MMSE}} = \frac{\beta_{iik}}{\zeta_{ik}}\mathbf{z}_{ik}, \tag{4.8}$$

where $\hat{\mathbf{g}}_{iik}^{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M)$ and the MSE of MMSE estimator is given by

$$\eta_{ik}^{\text{MMSE}} = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_{iik}^{\text{MMSE}} - \mathbf{g}_{iik}\|^2] = \beta_{iik}\left(1 - \frac{\beta_{iik}}{\zeta_{ik}}\right). \tag{4.9}$$

Remark 14. *Due to pilot contamination, as $q \to \infty$, $\eta_{ik}^{mmse} \to \beta_{iik}\left(1 - \frac{\beta_{iik}}{\sum_{l=1}^{L}\beta_{ilk}}\right)$.*

Remark 15. *If $\boldsymbol{g}_{iik}$ is an i.i.d. complex Gaussian vector, then, we see that $\hat{\boldsymbol{g}}_{ilk}^{MMSE} = \frac{\beta_{ilk}}{\beta_{iik}}\hat{\boldsymbol{g}}_{iik}^{MMSE}, \forall l$, meaning that the channel estimates are parallel vectors that only differ by the scaling factor, $\frac{\beta_{ilk}}{\beta_{iik}}$, and therefore, showing the BS's inability to separate users transmitting the same pilot sequence within other cells.*

From (4.8) it is possible to see that

$$\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} = \frac{\beta_{ilk}}{\beta_{iik}}\hat{\mathbf{g}}_{iik}^{\text{MMSE}}. \tag{4.10}$$

The channel estimate, $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}}$, can be expressed as

$$\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} = \tilde{\mathbf{g}}_{ilk}^{\text{MMSE}} + \mathbf{g}_{ilk}, \tag{4.11}$$

where $\tilde{\mathbf{g}}_{ilk}^{\text{MMSE}}$ is the estimation error. Therefore, we have

$$\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} \sim \mathcal{CN}\left(\mathbf{0}_M, \frac{\beta_{ilk}^2}{\zeta_{ik}}\mathbf{I}_M\right), \tag{4.12}$$

and

$$\tilde{\mathbf{g}}_{ilk}^{\text{MMSE}} \sim \mathcal{CN}\left(\mathbf{0}_M, \beta_{ilk}\left(1 - \frac{\beta_{ilk}}{\zeta_{ik}}\right)\mathbf{I}_M\right). \tag{4.13}$$

For MMSE channel estimation, the channel estimate $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}}$ is known to be independent of the channel estimation error $\tilde{\mathbf{g}}_{ilk}^{\text{MMSE}}$ [126].

4.4 Proposed Channel Estimator

In this section, we employ the ML method to estimate the parameter ζ_{ik} [126]. Applying the ML method to $f(\mathbf{z}_{ik}; \zeta_{ik}) \sim \mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$ we find the following estimator for ζ_{ik} given the observation $\mathbf{z}_{ik}$

$$\hat{\zeta_{ik}} = \frac{\|\mathbf{z}_{ik}\|^2}{M}. \tag{4.14}$$

This estimator exhibits a central Chi-square distribution with $2M$ degrees of freedom. It has $\mathbb{E}[\hat{\zeta_{ik}}] = \zeta_{ik}$, which shows that the ML estimator is unbiased, and $\text{var}\{\hat{\zeta_{ik}}\} = \zeta_{ik}^2/M$. In order to assess the efficiency of the estimator we derive the Cramér-Rao bound as [126]

$$\text{var}(\hat{\zeta_{ik}}) \geq \frac{\zeta_{ik}^2}{M}. \tag{4.15}$$

Therefore, the ML estimator derived for ζ_{ik} is the MVUE, *i.e.*, it is an unbiased estimator that has lower variance than any other unbiased estimator for all possible values of the parameter [126].

This simple and effective estimator is derived based on the observation that the MMSE estimator does not need to know the individual large-scale fading coefficients, $\{\beta_{ilk}\}$, as assumed in the existing literature, but that just ζ_{ik} suffices. The proposed estimator for ζ_{ik} renders the acquisition of inter-cell large-scale fading coefficients unnecessary. The task of gaining knowledge of those coefficients may be unjustifiable in practice due to the excessive overhead and additional latency imposed on the system. For instance, in case there are L cells serving K users in each one of them, each BS needs to acquire $(L-1)K$ inter-cell large-scale coefficients.

Swapping ζ_{ik} with $\hat{\zeta_{ik}}$ in (4.8) produces the proposed channel estimator, which is defined by

$$\hat{\mathbf{g}}_{iik}^{\text{prop}} = M\beta_{iik}\frac{\mathbf{z}_{ik}}{\|\mathbf{z}_{ik}\|^2}. \tag{4.16}$$

This estimator approaches the ideal MMSE estimator asymptotically with respect to M and has $\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}] = \mathbf{0}_M$ and covariance matrix given by

$$\begin{aligned}\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}(\hat{\mathbf{g}}_{iik}^{\text{prop}})^H] &= \left(\frac{M}{M-1}\frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\ &= \left(\frac{\beta_{iik}^2}{\zeta_{ik}} + \frac{1}{(M-1)}\frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\ &= \left(\frac{\beta_{iik}^2}{\zeta_{ik}} + \epsilon_{ik}\right)\mathbf{I}_M,\end{aligned} \tag{4.17}$$

where $\epsilon_{ik} = \frac{1}{(M-1)}\frac{\beta_{iik}^2}{\zeta_{ik}}$. The mean of $\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}]$ can be found by using the symmetry property of the distribution of $z_{ikm} \sim \mathcal{CN}(0, \zeta_{ik})$, hence we conclude that $\mathbb{E}\left[\hat{g}_{iikm}^{\text{prop}}\right] = 0$, once $\mathbb{E}\left[\frac{z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right] = \mathbb{E}\left[\frac{-z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right], \forall m$.

As can be seen by analyzing (4.17), as $M \to \infty$, $\text{Cov}[\hat{\mathbf{g}}_{iik}^{\text{prop}}] \to \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M$, which is the covariance matrix of the MMSE estimator. An approximation to the MSE per antenna of this estimator is given by

$$\eta_{ik}^{\text{prop}} = \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_{iik}^{\text{prop}} - \mathbf{g}_{iik}\|^2] \approx \beta_{iik}\left[1 - \frac{(M-2)\beta_{iik}}{(M-1)\zeta_{ik}}\right]. \tag{4.18}$$

The approximate MSE in (4.18) for the proposed estimator, decreases with increasing transmitting power q, increasing M or decreasing β_{iik}, which means smaller interference level from other cells, *i.e.*, smaller pilot contamination.

Remark 16. *Due to pilot contamination, as $q \rightarrow \infty$ and $M \rightarrow \infty$,* $\eta_{ik}^{prop} \rightarrow \beta_{iik}\left(1 - \frac{\beta_{iik}}{\sum_{l=1}^{L}\beta_{ilk}}\right)$.

Remark 16 clearly shows that the MSE of the proposed estimator tends to that of the MMSE estimator when both q and $M \rightarrow \infty$. The proof for the approximation of the MSE is given in Appendix F.

Remark 17. *If $\boldsymbol{g}_{iik}$ is an i.i.d. complex Gaussian vector, then, we see that $\hat{\boldsymbol{g}}_{ilk}^{prop.} = \frac{\beta_{ilk}}{\beta_{iik}}\hat{\boldsymbol{g}}_{iik}^{prop.}, \forall l$, meaning that, like what happens with the MMSE channel estimator, the channel estimates are parallel vectors that only differ by the scaling factor, $\frac{\beta_{ilk}}{\beta_{iik}}$, and therefore, showing the BS's inability to separate users transmitting the same pilot sequence within other cells.*

The difference between the closed-form, given by (9) in [143], and the approximated MSE expressions is defined by

$$\eta_{ik}^{\text{prop(closed-form)}} - \eta_{ik}^{\text{prop(approx.)}} = 2\beta_{iik}\left\{\frac{\beta_{iik}}{\zeta_{ik}} - \theta_{ik}\right\}, \tag{4.19}$$

where θ_{ik} is defined in [143].

Remark 18. *As both q and $M \rightarrow \infty$, $\theta_{ik} \rightarrow \frac{\beta_{iik}}{\zeta_{ik}}$ and then, $\eta_{ik}^{prop(closed\text{-}form)} - \eta_{ik}^{prop(approx.)} \rightarrow 0$.*

We find Remark 18 by using Remark 16 and equaling the closed-form and approximated MSE expressions. This remark shows that the difference between the closed-form and the approximated MSE expressions decreases, tending to 0, as both UL power, q, and the number of receiving antennas, M, increase.

Remark 19. *The average normalized squared Euclidean distance between $\hat{\boldsymbol{g}}_{iik}^{prop}$ and $\hat{\boldsymbol{g}}_{iik}^{MMSE}$ is given by*

$$\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}} - \hat{\mathbf{g}}_{iik}^{\text{MMSE}}\|^2\right] = \frac{1}{M-1}\frac{\beta_{iik}^2}{\zeta_{ik}} = \epsilon_{ik}. \tag{4.20}$$

The proof of (4.20) is given in Appendix G. From (4.5) and (4.20), it is easily noticeable that the average distance decreases with increasing M, decreasing q, increasing $\beta_{ilk}, i \neq l$, and decreasing β_{iik}.

Remark 20. *The channel estimation error,* $\tilde{\boldsymbol{g}}_{iik}^{prop} = \hat{\boldsymbol{g}}_{iik}^{prop} - \boldsymbol{g}_{iik}$*, is correlated with the channel estimate and uncorrelated with the pilot signal,* $\boldsymbol{z}_{ik}$*,*

$$\frac{1}{M}\mathbb{E}\left[(\tilde{\mathbf{g}}_{iik}^{\text{prop}})^H \hat{\mathbf{g}}_{iik}^{\text{prop}}\right] = \epsilon_{ik}, \tag{4.21}$$

$$\frac{1}{M}\mathbb{E}\left[(\tilde{\mathbf{g}}_{iik}^{\text{prop}})^H \mathbf{z}_{ik}\right] = 0, \tag{4.22}$$

The proof of (4.21) follows the same rationale used to find the third term of (F.10) in Appendix F. It is clear that (4.21), as (4.20), decreases with increasing M, decreasing q, increasing $\beta_{ilk}, i \neq l$, and decreasing β_{iik}. The estimation error, $\tilde{\mathbf{g}}_{iik}^{\text{prop}}$, has the following mean vector and covariance matrix,

$$\mathbb{E}[\tilde{\mathbf{g}}_{iik}^{\text{prop}}] = \mathbf{0}_M, \tag{4.23}$$

$$\begin{aligned}\text{Cov}\,(\tilde{\mathbf{g}}_{iik}^{\text{prop}}) &= \mathbb{E}[\tilde{\mathbf{g}}_{iik}^{\text{prop}}(\tilde{\mathbf{g}}_{iik}^{\text{prop}})^H] \\ &\approx \left(\beta_{iik} + \epsilon_{ik} - \frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\ &= \beta_{iik}\left[1 - \frac{(M-2)\beta_{iik}}{(M-1)\zeta_{ik}}\right] \\ &= \left[\beta_{iik}\left(1 - \frac{\beta_{iik}}{\zeta_{ik}}\right) + \epsilon_{ik}\right]\mathbf{I}_M.\end{aligned} \tag{4.24}$$

Remark 21. *As* $M \to \infty$ *the covariance matrix of the estimation error,* $\tilde{\boldsymbol{g}}_{iik}^{prop}$*, tends to* $\mathbb{E}[\tilde{\boldsymbol{g}}_{iik}^{MMSE}(\tilde{\boldsymbol{g}}_{iik}^{MMSE})^H] = \left[\beta_{iik}\left(1 - \frac{\beta_{iik}}{\zeta_{ik}}\right)\right]\boldsymbol{I}_M$.

4.5 Simulation Results

In this section, we compare the performance of the proposed channel estimator with that of the MMSE and LS estimators. We adopt a typical multi-cell structure as the one shown in Fig. 4.1 with $L = 7$ cells (one central cell surrounded by 6 other cells), $K = 10$ users in each cell, frequency reuse factor of 1 and $N = K$ pilot symbols. We consider two different types of setups for $\{\beta_{ilk}\}$, one with fixed values and other with random values. For the fixed case, we set $\beta_{iik} = 1$ and $\beta_{ilk} = a, \forall\, l \neq i$, where a represents the cross-cell interference level. The value selected for a in the fixed case is 0.05 and it is chosen so that there is moderate cross-cell interference level from users being served by other BSs, *i.e.*, not being served by the central cell. For the random case, users in each cell are uniformly distributed within a ring with radii $d_0 = 100$ m and $d_1 = 1000$ m respectively. The large-scale fading coefficients $\{\beta_{ilk}\}$ are independently generated by

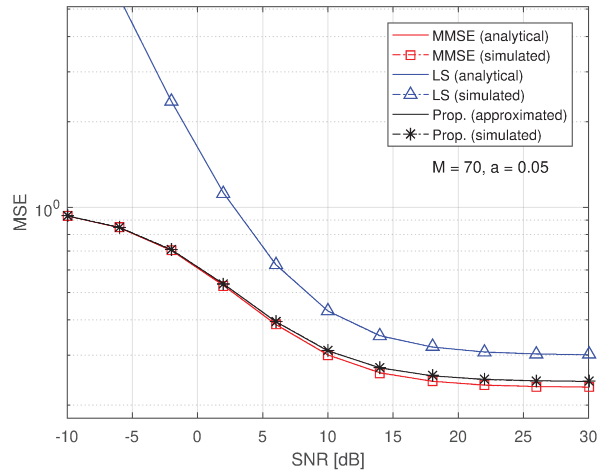

Figure 4.3 – Channel Estimation MSE versus UL pilot power.

$\beta_{ilk} = \psi / \left(\frac{d_{ilk}}{d_0}\right)^v$, where $v = 3.8$, $10 \log_{10}(\psi) \sim \mathcal{N}(0, \sigma^2_{\text{shadow,dB}})$ with $\sigma_{\text{shadow,dB}} = 8$, and d_{ilk} is the distance of the k-th user in the l-th cell to the i-th BS. Both, the path loss exponent, v, and the standard deviation of the log-normal shadowing, $\sigma_{\text{shadow,dB}}$, are common values for outdoor shadowed urban cellular radio environments [103, 116].

The results in Fig. 4.3 show MSE versus SNR (UL pilot power q) performances for $a = 0.05$ and $M = 70$. As can be seen, analytical, approximated and simulation MSEs match for all estimators. With the increase of SNR, MSEs of all the estimation methods decrease. There are MSE floors for all the three estimators due to pilot contamination (see Remarks 12, 14 and 16). At low SNR, the MSE of the proposed estimator is very close to that of the ideal MMSE estimator. On the other hand, as can be noticed, with the increase of the SNR, the gap between the ideal MMSE estimator and the proposed one increases (see Remark 19).

In Fig. 4.4, we compare MSE versus the number of BS antennas M under the

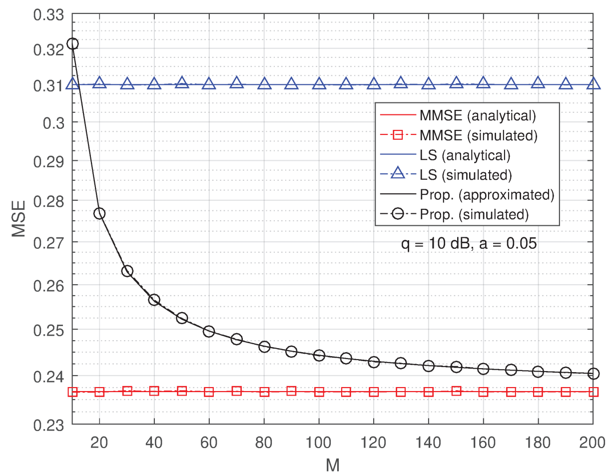

Figure 4.4 – MSE performance versus number of BS collocated antennas, M.

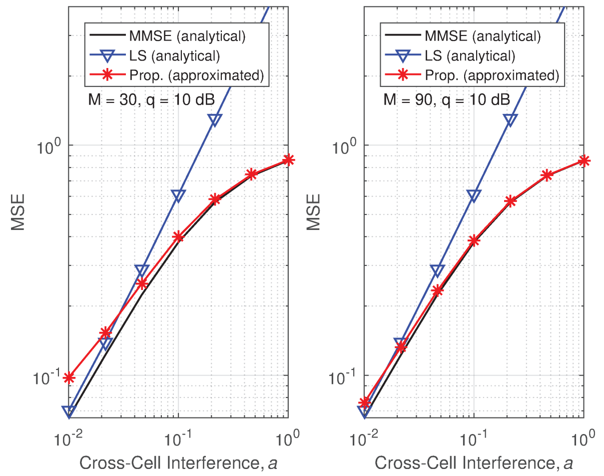

Figure 4.5 – Channel estimation MSE versus cross-cell interference level.

setting of $a = 0.05$ and TX SNR $q = 10$ dB. With the increase of M, the MSE of the proposed estimator approaches that of the ideal MMSE, while the MSE of LS estimator does not change. Due to numerical issues, the closed-form MSE expression presented in [143] does not produce values for $M > 85$. During our simulations, comparing the closed-form expression given by (9) in [143] and the approximated MSE expression given by (4.18), we noticed that the $\Gamma(2M)$ function in the numerator of (9) in [143] grows without bound, reaching values that are greater than the largest possible finite floating-point number represented by the Institute of Electrical and Electronics Engineers (IEEE) double precision format, *i.e.*, 1.7977e+308 [144], for values of M greater than 85. A double precision variable goes to +Inf after the largest possible number [144]. On the other hand, as can be seen in Fig. 4.4, the approximate analytical MSE expression (4.18) does not present the same problem and, therefore, can be used to evaluate the MSE for any number of antennas, M without any numerical issue.

In Fig. 4.5, we compare MSE performance with respect to various levels of

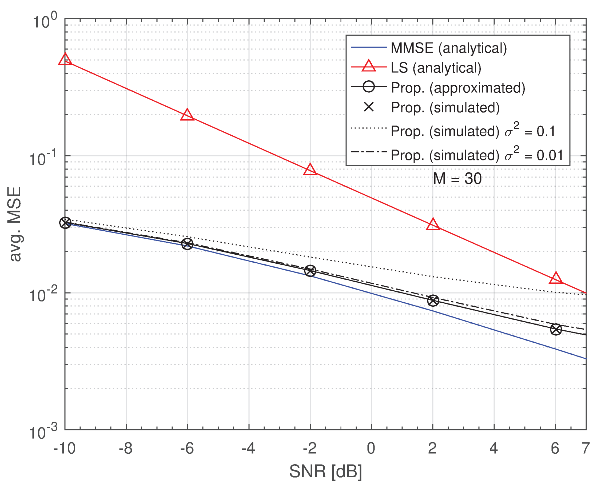

Figure 4.6 – Average channel estimation MSE under random $\{\beta_{ilk}\}$.

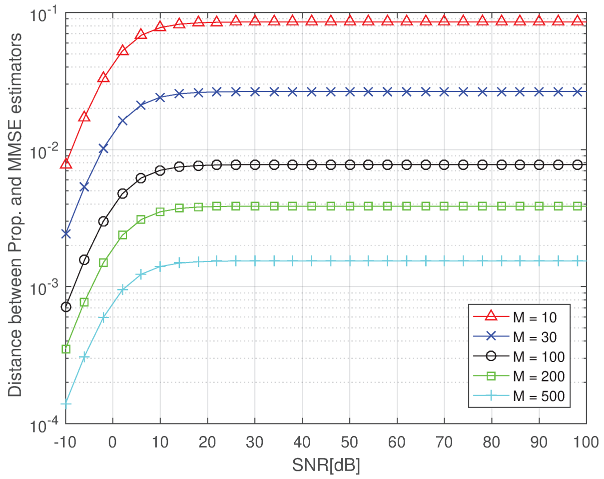

Figure 4.7 – Distance between proposed and MMSE estimators (Remark 19).

cross-cell interference, a, with $q = 10$ dB and two different number of antennas, $M = 30$ and $M = 90$. We can see that when a increases (the effect of pilot contamination increases) the estimation performance degrades. At a low cross-cell interference level, LS presents a slightly better MSE when compared to the proposed estimator. This difference disappears as M increases, as can be noticed in the plot with $M = 90$. As the interference level increases, the proposed method outperforms the LS estimator substantially and approaches the ideal MMSE performance (see Remark 19).

In Fig. 4.6, we evaluate the MSE performance under random large-scale fading coefficients $\{\beta_{ilk}\}$ with $M = 30$. The results are obtained by averaging MSEs over 10^4 realizations of $\{\beta_{ilk}\}$. As can be observed, simulation MSE matches with the analytical MSE. Additionally, the sensitivity of the proposed estimator against inaccuracy of β_{iik} by using an estimate $\hat{\beta}_{iik} = \beta_{iik}(1 + \mathcal{N}(0, \sigma^2))$ is investigated. The performance degradation for $\sigma^2 = 0.1$ is noticeable at high SNR but for $\sigma^2 = 0.01$ it is insignificant. The proposed estimator still outperforms the LS estimator significantly.

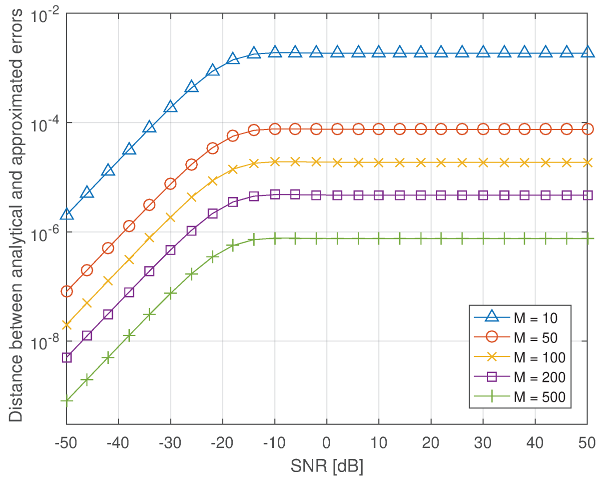

Figure 4.8 – Absolute distance between closed-form and approximated MSE expressions.

In Fig. 4.7, we compare the distance between the proposed and MMSE channel estimators for different number of antennas, M, with $a = 0.05$. As the Remark 19 states, the distance is small at low SNR, increasing with SNR until a ceiling is reached. As can be also noticed, the ceiling value decreases with the number of antennas, M.

In Fig. 4.8, we compare the absolute distance between the approximated MSE expression presented in (4.18) and the analytical (closed-form) MSE expression presented in [143] for various SNR and M values with $a = 0.05$. The distance between the MSE expressions is small at low SNR, increasing with SNR until a ceiling value is reached. As can be noticed, the ceiling value decreases with the number of antennas, M. For $M = 50$ the ceiling distance is smaller than 10^{-4}, showing that the approximated MSE expression can replace the one presented in [143].

4.6 Conclusions

In this chapter, we have introduced a simple and practical channel estimator for massive MIMO TDD systems with pilot contamination in a flat channel environment. The proposed estimator replaces the combined interference plus noise power term in the ideal MMSE estimator with a Maximum Likelihood estimator for that term. Moreover, the proposed estimator presents MSE results that are very close to that of the ideal MMSE estimator without requiring previous knowledge of noise and interference statistics. Additionally, we have derived an approximate analytical MSE expression for the proposed estimator which can be useful in system design and performance evaluation. We have also shown that the MSE expression presented here asymptotically approaches that of the MMSE estimator. Finally, the simpler approximate analytical MSE expression presented here can be used instead of the more complex and susceptible to numerical issues one presented in [143].

5 On the Distribution of an Effective Channel Estimator for Multi-cell Massive MIMO systems

5.1 Introduction

The estimation of the inter-cell large scale fading coefficients might be unjustified in practice due to the excessive overhead it imposes on the system [79]. For instance, in case there are L cells serving K users in each one of them, each BS needs to acquire $K(L-1)$ inter-cell large-scale fading coefficients. In chapter 4 we proposed a channel estimator for multi-cell multi-user systems that does not require the estimation of the inter-cell large-scale fading coefficients. In this chapter, we find expressions for the PDF, CDF and central moments of the channel estimator proposed in the previous chapter.

5.2 UL Signal Model

As in chapter 4, in this chapter, we also consider a multi-cell system with L cells where each cell has at its center a BS with M antennas and K randomly located single-antennas users in each one of the cells. We assume frequency-flat fading channels with the overall channel matrix being denoted by the $M \times K$ matrix $\mathbf{G}_{il}$, where its k-th column, $\mathbf{g}_{ilk} = [g_{ilk1}, \cdots, g_{ilkM}]^{\mathrm{T}}$, represents the channel gain from the k-th user in the l-th cell to all the M antennas at the i-th BS, with $g_{ilkm} = \sqrt{\beta_{ilk}} h_{ilkm} \sim \mathcal{CN}(0, \beta_{ilk})$. As before, we consider the set of large-scale coefficients, $\{\beta_{ilk}\}$, as being deterministic during the estimation phase.

5.3 UL Training

We assume that the users of all cells use the same set of pilot sequences at the same time (*i.e.*, all users' transmissions are aligned to the BS UL) and that the pilot reuse factor is equal to one, the most aggressive one. The pilot sequences of K users are represented by a $\tau \times K$ matrix Φ with the orthogonality property, $\Phi^H \Phi = \mathbf{I}_K$, where $K \leq \tau$. The received pilot sequences at the i-th BS are represented by a $M \times \tau$ matrix, $\mathbf{Y}_i$, defined as

$$\mathbf{Y}_i = \sqrt{\rho} \sum_{l=1}^{L} \mathbf{G}_{il} \Phi^H + \mathbf{N}_i, \tag{5.1}$$

where ρ is the average pilot transmit power of each user and $\mathbf{N}_i$ is a $M \times \tau$ matrix with i.i.d. elements following the distribution $\mathcal{CN}(0,1)$. Let ϕ_k denote the k-th column of Φ^H. Hence, a sufficient statistic for the estimation of the channel vectors, $\mathbf{g}_{iik}$, at the i-th BS is given by

$$\mathbf{z}_{ik} = \frac{1}{\sqrt{\rho}}\mathbf{Y}_i\phi_k = \sum_{l=1}^{L}\mathbf{G}_{il}\Phi^H\phi_k + \frac{1}{\sqrt{\rho}}\mathbf{N}_i\phi_k = \sum_{l=1}^{L}\mathbf{g}_{ilk} + \mathbf{w}_{ik}, \tag{5.2}$$

where $\mathbf{w}_{ik} = \frac{1}{\sqrt{\rho}}\mathbf{N}_i\phi_k$ has distribution $\mathcal{CN}(\mathbf{0}_M, \frac{1}{\rho}\mathbf{I}_M)$ and $\mathbf{z}_{ik}$ follows the distribution $\mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$ where $\zeta_{ik} = \sum_{l=1}^{L}\beta_{ilk} + \frac{1}{\rho}$. During detecton phase, the i-th BS has to estimate the channels of its users, *i.e.*, $\mathbf{g}_{ilk}, \forall k$.

5.4 Effective Channel Estimator and its distribution

A very simple but yet effective channel estimator for the multi-cell case is defined by (4.16), which asymptotically approaches the MMSE estimator as $M \to \infty$ [79]. Since $\mathbf{z}_{ik} \sim \mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$, then $\|\mathbf{z}_{ik}\|^2 \sim \Gamma(M, \zeta_{ik})$, therefore, the distribution of the elements, $\hat{g}_{iikm}^{\text{prop.}} = \frac{M\beta_{iik}z_{ikm}}{\|\mathbf{z}_{ik}\|^2}$, of the channel estimator, $\hat{\mathbf{g}}_{iik}^{\text{prop.}}$, is the ratio between a Circularly-symmetric normal distribution and a Gamma distribution. As $z_{ikm}\forall m$ are i.i.d. complex normal random variables, then it is clear that the distribution of $\hat{g}_{iikm}^{\text{prop.}}$ in $\mathbb{C} = \mathbb{R}^2$ is rotation invariant. Therefore, it suffices to find the distribution of $|\hat{g}_{iikm}^{\text{prop.}}|^2$, which is the same as that of $R = \frac{bX}{(X+Y)^2} \geq 0$, where $b = M^2\beta_{iik}^2$, X and Y are independent random variables exhibiting the distributions $\Gamma(1, \zeta_{ik})$ and $\Gamma(M-1, \zeta_{ik})$, respectively. Next, we consider the transformation of random variables from (X, Y) to (R, S), where $S = X$ and R as defined earlier. After finding the joint distribution of (R, S), $f_{R,S}(r, s)$, then the PDF of R is defined by $\int_0^\infty f_{R,S}(r,s)ds$. After applying the substitution $v = \sqrt{rs}$ to the previous integral, we find

$$f_R(r) = \frac{(M\beta_{iik})^{2M}}{\zeta_{ik}^M(M-2)!r^{M+1}}\int_0^1 v^M(1-v)^{M-2}e^{-\frac{(M\beta_{iik})^2 v}{\zeta_{ik}r}}dv, \tag{5.3}$$

which by using an integral solver [146] results in

$$\begin{aligned} f_R(r) &= \frac{1}{\zeta_{ik}r^2}\sqrt{\frac{(M\beta_{iik})^2\pi}{\zeta_{ik}r}}\frac{e^{-\frac{(M\beta_{iik})^2}{2\zeta_{ik}r}}}{2} \\ \times&\left\{\left[2\zeta_{ik}Mr - (M\beta_{iik})^2\right]I_{M-\frac{1}{2}}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik}r}\right) + (M\beta_{iik})^2 I_{M+\frac{1}{2}}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik}r}\right)\right\}, \end{aligned} \tag{5.4}$$

where $I_n(z)$ is the modified Bessel function of the first kind. From (5.3) and using the Fubini theorem [145], it is easy to find all the central moments of $R = |\hat{g}_{iikm}^{\text{prop.}}|^2$, which are defined as

$$\mathbb{E}\left[R^k\right] = \mathbb{E}\left[|\hat{g}_{iikm}^{\text{prop.}}|^{2k}\right] = \frac{(M\beta_{iik})^{2k}k!(M-k-1)!}{\zeta_{ik}^k(M+k-1)!}, k \in \mathbb{Z}. \tag{5.5}$$

The variance of $|\hat{g}_{iikm}^{\text{prop.}}|^2$ is defined as

$$\text{var}\left(|\hat{g}_{iikm}^{\text{prop.}}|^2\right) = \frac{(M\beta_{iik})^4[2+M(M-1)]}{\zeta_{ik}^2 M^2(M+1)(M-1)^2(M-2)}, \tag{5.6}$$

The covariance between $|\hat{g}_{iikm}^{\text{prop.}}|^2$ and $|\hat{g}_{iikn}^{\text{prop.}}|^2$ is given by

$$\text{cov}\left(|\hat{g}_{iikm}^{\text{prop.}}|^2, |\hat{g}_{iikn}^{\text{prop.}}|^2\right) = \frac{2(M\beta_{iik})^4}{\zeta_{ik}^2 M^2(M+1)(M-1)^2(M-2)}, \tag{5.7}$$

which tends to 0 when $M \to \infty$. By using the symmetry property of the distribution of $z_{ikm} \sim \mathcal{CN}(0, \zeta_{ik})$, we conclude that $\mathbb{E}\left[\hat{g}_{iikm}^{\text{prop.}}\right] = 0$, once $\mathbb{E}\left[\frac{z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right] = \mathbb{E}\left[\frac{-z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right]$. Next, by using the symmetry property of z_{ikm} again, we find that $\hat{g}_{iikm}^{\text{prop.}}$ and $\hat{g}_{iikn}^{\text{prop.}}$, when $m \neq n$, are uncorrelated, *i.e.*, $\text{cov}(\hat{g}_{iikm}^{\text{prop.}}, \hat{g}_{iikn}^{\text{prop.}}) = \mathbb{E}\left[(\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}\right]$, once $\mathbb{E}\left[\hat{g}_{iikm}^{\text{prop.}}\right] = \mathbb{E}\left[\hat{g}_{iikn}^{\text{prop.}}\right] = 0$. First, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}\left[|(\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}|\right] &\leq \sqrt{\mathbb{E}\left[|\hat{g}_{iikm}^{\text{prop.}}|^2\right] \mathbb{E}\left[|\hat{g}_{iikn}^{\text{prop.}}|^2\right]} \\
&= \mathbb{E}\left[|\hat{g}_{iikm}^{\text{prop.}}|^2\right] = \mathbb{E}\left[|\hat{g}_{iikn}^{\text{prop.}}|^2\right] \\
&= \frac{M\beta_{iik}^2}{\zeta_{ik}(M-1)} < \infty, m \neq n,
\end{aligned} \tag{5.8}$$

hence, $\mathbb{E}\left[(\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}\right]$ exists and is finite. Therefore and because the joint distribution of the pair $((-\hat{g}_{iikm}^{\text{prop.}})^*, \hat{g}_{iikn}^{\text{prop.}})$ is the same as that of $((\hat{g}_{iikm}^{\text{prop.}})^*, \hat{g}_{iikn}^{\text{prop.}})$, we conclude that $\mathbb{E}\left[(\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}\right] = \mathbb{E}\left[(-\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}\right] = 0$. Therefore, the elements of $\hat{\mathbf{g}}_{iik}^{\text{prop.}}$ are uncorrelated.

The CDF of $R = |\hat{g}_{iikm}^{\text{prop.}}|^2$ is defined as

$$F_R(r) = 1 - \sqrt{\frac{(M\beta_{iik})^2 \pi}{\zeta_{ik} r}} e^{-\frac{(M\beta_{iik})^2}{2\zeta_{ik} r}} I_{M-\frac{1}{2}}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik} r}\right), M > 1. \tag{5.9}$$

Next, we find the PDF and CDF of $U = \Re\{\hat{g}_{iikm}^{\text{prop.}}\} = \Im\{\hat{g}_{iikm}^{\text{prop.}}\}$. First we recall that $\hat{g}_{iikm}^{\text{prop.}}$ can be expressed as $\sqrt{R}C + j\sqrt{R}S$, where $C = \cos(\Theta)$, $S = \sin(\Theta)$, and Θ is a random variable independent of R and uniformly distributed in the interval $[0, 2\pi]$ or, by symmetry, in $[0, \pi]$. The PDF of C, is given by $f_C(c) = \frac{1}{\pi\sqrt{1-c^2}}, -1 < c < 1$. Considering the transformation of random variables from (R, C) to (U, W), where $U = \sqrt{R}C$ and $W = R$, we have the following for the PDF of U

$$f_U(u) = \frac{1}{\pi} \int_{u^2}^{\infty} \frac{1}{\sqrt{w-u^2}} f_R(w) dw, \tag{5.10}$$

for all real u, where $f_R(r)$ is given by (5.3). Using the Fubini theorem and an integral solver [146] for the calculation of the iterated double integral, we get

$$f_U(u) = \frac{(M\beta_{iik})^2 e^{-\frac{(M\beta_{iik})^2}{2\zeta_{ik} u^2}}}{2\zeta_{ik}|u|^3} \left[I_{M-1}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik} u^2}\right) - I_M\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik} u^2}\right)\right]. \tag{5.11}$$

By recalling that $U = \sqrt{R}\cos(\Theta)$ and that R and Θ are independent, we know that all odd moments of U vanish due to the fact that $\mathbb{E}\left[(\cos(\Theta))^{2k+1}\right] = 0, k \in \mathbb{Z}$ and that all even moments of U are given by

$$\mathbb{E}\left[U^{2k}\right] = \mathbb{E}\left[R^k\right]\mathbb{E}\left[(\cos(\Theta))^{2k}\right] = \frac{(M\beta_{iik})^{2k}}{\zeta_{ik}^k}\frac{(M-k-1)!}{(M+k-1)!}\frac{\Gamma(k+\frac{1}{2})}{\sqrt{\pi}}. \tag{5.12}$$

The covariance between $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ and $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ is define as

$$\begin{aligned}\operatorname{cov}\left(\Re\{\hat{g}_{iikm}^{\text{prop.}}\}, \Im\{\hat{g}_{iikm}^{\text{prop.}}\}\right) &= \mathbb{E}\left[\Re\{\hat{g}_{iikm}^{\text{prop.}}\}\Im\{\hat{g}_{iikm}^{\text{prop.}}\}\right] \\ &= \mathbb{E}\left[\sqrt{R}C\sqrt{R}S\right] = \mathbb{E}\left[RCS\right] = \mathbb{E}\left[R\right]\mathbb{E}\left[CS\right] = 0,\end{aligned} \tag{5.13}$$

once $\mathbb{E}\left[CS\right] = \frac{1}{\pi}\int_0^\pi \cos(\Theta)\sin(\Theta)d\Theta = 0$.

By using (5.12), the Kurtosis of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\} = \Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ is defined as

$$\operatorname{Kurt}\left[\Re\{\hat{g}_{iikm}^{\text{prop.}}\}\right] = \frac{\mathbb{E}[U^4]}{(\mathbb{E}[U^2])^2} = \frac{3M(M-1)}{(M+1)(M-2)} \overset{M\to\infty}{\approx} 3, M > 2, \tag{5.14}$$

which approaches 3 as the number of antennas, M, increases. Therefore, as $M \to \infty$, the distribution of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ and $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ becomes a mesokurtic distribution, *i.e.*, this distribution has tails shaped the same way as the normal distribution as M increases. Hence, as $M \to \infty$ the distributions of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ and $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ tend to that of the normal distribution with mean equal to 0 and variance equal to $\frac{M\beta_{iik}^2}{2(M-1)\zeta_{ik}}$.

As $\operatorname{cov}\left(\hat{g}_{iikm}^{\text{prop.}}, \hat{g}_{iikn}^{\text{prop.}}\right) = 0, m \neq n$, $\operatorname{cov}\left(\Re\{\hat{g}_{iikm}^{\text{prop.}}\}, \Im\{\hat{g}_{iikm}^{\text{prop.}}\}\right) = 0, \forall m$, and $\operatorname{Kurt}\left[\Re\{\hat{g}_{iikm}^{\text{prop.}}\}\right] \to 3$ as M increases, then we conclude that the distribution of random vector $\hat{\mathbf{g}}_{iik}^{\text{prop.}}$, approaches that of a circularly-symmetric complex normal vector, $\hat{\mathbf{g}}_{iik}^{\text{prop.}} \sim \mathcal{CN}\left(\mathbf{0}_M, \frac{M\beta_{iik}^2}{(M-1)\zeta_{ik}}\mathbf{I}_M\right)$ as M increases. The CDF of U, $F_U(u) = 2\int_0^u f_U(t)dt$, results in a quite complicated expression involving Hypergeometric functions, *i.e.*, it does not have a neat closed form expression, however, as we just discussed, it can be approximated by the CDF of a normal random variable as M increases.

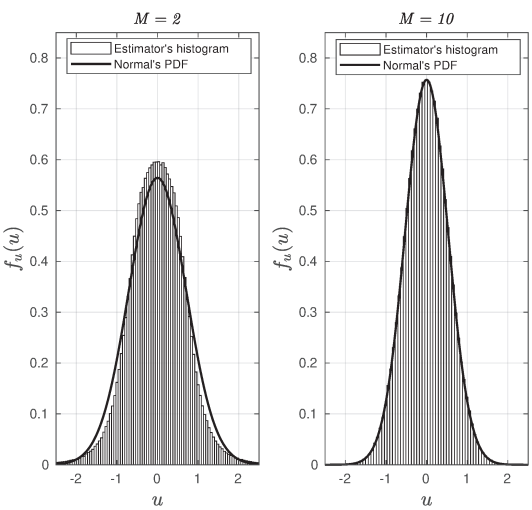

Figure 5.1 – Proposed estimator's histogram versus Normal's PDF.

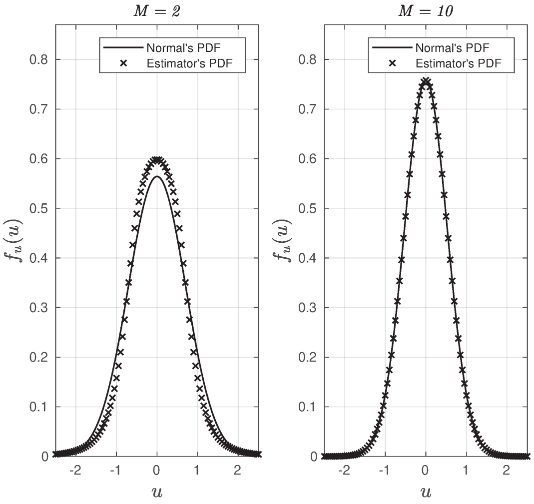

Figure 5.2 – Eq. (5.11) versus the Normal's PDF.

5.5 Simulation Results

In this section, we present some results showing that the distribution of the proposed estimator asymptotically approaches that of the circularly-symmetric complex normal distribution. Figures 5.1 to 5.4 show several results comparing the distribution of the real (or imaginary) part, $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$, of the proposed estimator and the Normal distribution for $\beta_{iik} = 1$ and $\zeta_{ik} = \beta_{iik} + \sum_{l=1,l\neq i}^{L} \beta_{ilk} + \frac{1}{\rho} = 2$ (*i.e.*, $\sum_{l=1,l\neq i}^{L} \beta_{ilk} + \frac{1}{\rho} = 1$) respectively. The large-scale coefficient of the device within the target cell, β_{iik}, was chosen so that it was greater than the other individual large-scale coefficients from interfering cells. The value of ζ_{ik} was chosen so that the inter-cell large-scale fading coefficients plus noise power term, $\sum_{l=1,l\neq i}^{L} \beta_{ilk} + \frac{1}{\rho}$, is comparable to β_{iik}.

Figure 5.1 compares the normalized histogram of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ with the PDF of the Normal distribution, $\mathcal{N}\left(0, \frac{M\beta_{iik}^2}{2(M-1)\zeta_{ik}}\right)$. The figure shows that for $M = 2$ the distributions

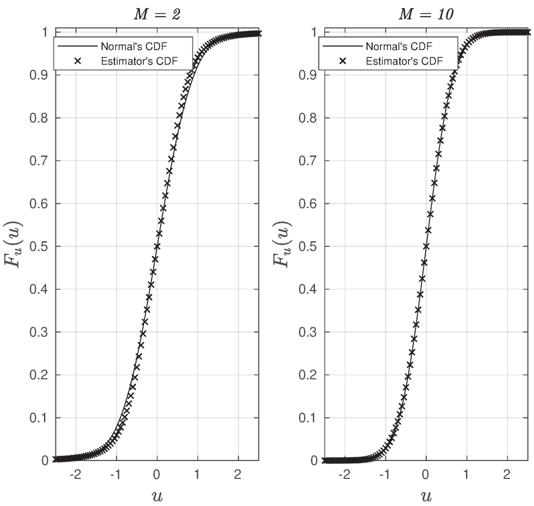

Figure 5.3 – Proposed estimator's CDF, $F_U(u) = 2\int_0^u f_U(t)dt$, versus the Normal's CDF.

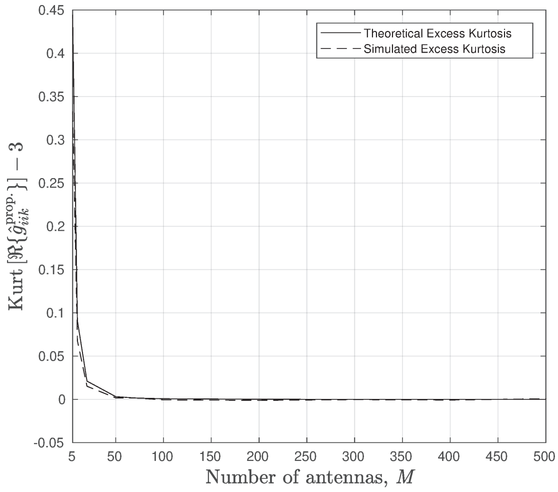

Figure 5.4 – Proposed estimator's excess Kurtosis for several number of antennas, M.

differ considerably, however, for $M = 10$ they become very close.

Figures 5.2 and 5.3 compare the PDF and CDF of U, with the PDF and CDF of the Normal distribution respectively, again the results show that they become quite close as M increases.

Figure 5.4 shows that the excess Kurtosis of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ tends to 0 as M increases, showing that, indeed, the distribution of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ asymptotically tends to that of the Normal distribution.

Figure 5.5 shows that the MAPE between $f_U(u)$, given by (5.11), and the Normal's PDF, $f_N(n)$, decreases as M increases.

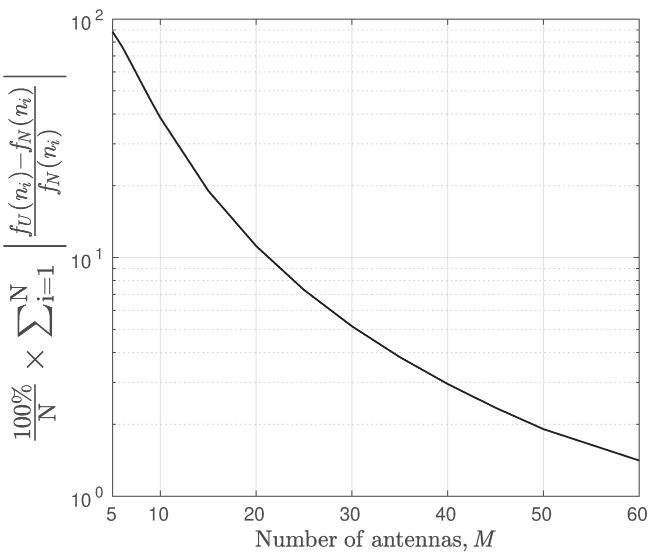

Figure 5.5 – MAPE between (5.11) and the Normal's PDF, $f_N(n)$ for several number of antennas, M.

5.6 Conclusions

In this chapter, we have derived expressions for the PDF, CDF and central moments of the channel estimator proposed in chapter 4. We have concluded that the distribution of the channel estimator asymptotically approaches the circularly-symmetric complex normal distribution as the number of antennas M increases. The results presented here show that the distribution of the estimator proposed in chapter 4 can be accurately approximated by the circularly-symmetric complex normal distribution for $M > 10$, which is the case for massive MIMO deployments [147].

6 Channel Estimation for SC-FDMA Multi-Cell Massive MIMO Systems Considering Pilot Contamination

6.1 Introduction

Massive Multiple-Input Multiple-Output, antenna systems potentially allow BS to operate with huge improvements in spectral and radiated energy efficiency, using relatively low complexity linear processing techniques. The improvement in SE is achieved by serving several terminals in the same time-frequency resource through spatial multiplexing, and the increase in energy efficiency is due to the array gain yielded by the large number of antennas available at the BS [4, 42].

The expected massive MIMO improvements assume that accurate channel estimations are available at both the receiver and transmitter for detection and precoding operations, respectively. Additionally, the reuse of frequencies and pilot sequences in cellular communications systems leads to degradation in the channel estimation performance. As mentioned in the literature, this inter-cellular interference is known as pilot contamination [42, 69].

Massive MIMO systems operating in TDD mode assume channel reciprocity between UL and DL in order to minimize pilot overhead, transmitting pilot reference signals only in the UL. In this scenario, the pilot overhead cost is only proportional to the number of terminals and improved channel estimation quality can be achieved due to the large number of antennas available at the BS [41, 43, 69]. A BS estimates channel states usually based on LS [76] or MMSE [129, 130] linear estimators. In addition, inter and intra-cell large-scale fading coefficients are assumed to be perfectly known at the BS when applying the MMSE estimator in a great number of works [102, 129, 130].

A countless number of works in the literature only deals with the problem of channel estimation in flat fading channels [41, 69, 76, 77, 102, 129–132, 143, 148–152], which is applicable only for OFDM based systems and does not reflect the true nature of communications channels in practice [103].

In this chapter, we study the channel estimation problem and its influences on the performance of SC-FDMA multi-cell massive MIMO TDD systems under the assumptions of time-varying frequency-selective fading channels and frequency/pilot-sequence

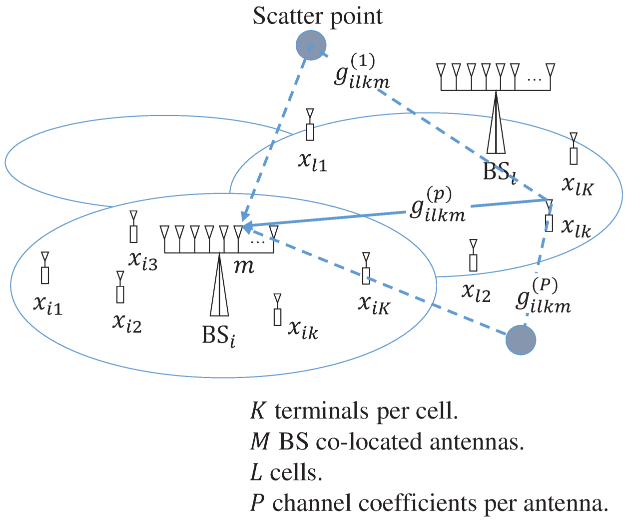

Figure 6.1 – Problem definition.

reuse. A significant disadvantage of systems employing OFDM is its high PAPR. Additionally, when a massive MIMO scenario is considered, the addition of FFT and IFFT blocks per antenna branch will considerably increase the implementation complexity of the transceivers. Therefore, a possible solution to these issues is the use of SC-FDMA systems, which inherently present lower PAPR and have relatively simpler implementation when compared to OFDM systems [153].

Our main contributions in this chapter are: (i) the proposal of an UL training scheme using ZC sequences to ease the channel estimation process; (ii) the use of the MVUE method to estimate the interference (*i.e.*, inter-cell large-scale fading coefficients) plus noise power term and replace it into the ideal MMSE channel estimator, resulting in an efficient, more practical and simpler channel estimator; (iii) the derivation of an approximate analytical MSE expression for the proposed estimator; and (iv) a thorough analysis of the proposed channel estimator and its approximate MSE expression. Numerical results demonstrate that the proposed channel estimator performs asymptotically as well as the ideal MMSE channel estimator as the number of antennas and channel paths increases.

The chapter is organized in the following way. Section 6.2 defines the problem at hand and briefly discusses two linear channel estimators, namely, LS and MMSE, which will be used in comparative analysis with the proposed channel estimator. Section 6.3 presents and analyses the proposed channel estimator and its approximate analytical MSE expression. The performance of the proposed estimator is assessed in Section 6.4. Section 6.5 then concludes the chapter.

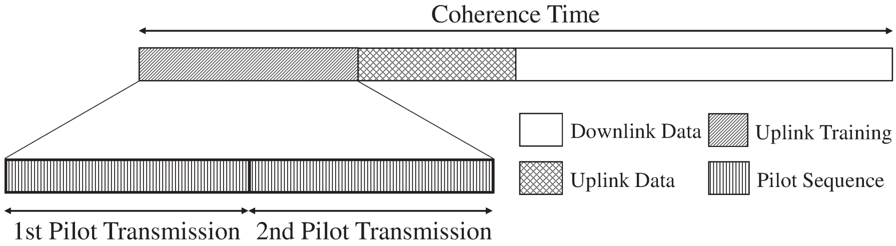

Figure 6.2 – TDD transmission protocol.

6.2 Problem Definition

In this chapter, we consider a multi-cell MU massive MIMO system with L cells, where a TDD protocol is employed. Each one of the cells has a BS at its center equipped with M co-located antennas and K single-antenna users randomly located within the cell, where $M \gg K$. All the K users in a cell are served at the same time/frequency resource. Fig. 6.1 illustrates the problem we are dealing with, where the solid arrow corresponds to the possible direct path and the dashed lines correspond to paths where the signal from the k-th user in the l-th cell reaches the m-th antenna at the i-th base station through reflections caused by objects such as buildings and mountains.

Based on the assumption of channel reciprocity, we adopt the TDD protocol depicted in Fig. 6.2 and proposed in [138]. Due to the reciprocity principle, only UL channels need to be estimated while DL channels are equal to the conjugate-transpose of the UL channels. It is important to note that the length of the TDD frames is limited by the channel coherence time [138, 139]. According to the adopted TDD protocol, first, all users in all cells transmit their UL training sequences synchronously. After that, the BSs use the training sequences to estimate the UL channels. Next, the users send UL data signals. Then, the BSs use the estimated channels to detect and decode UL data and generate precoding matrices employed to transmit DL data.

In a multi-cell scenario, the set of pilot sequences has to be re-utilized by users in neighboring cells once orthogonal pilot sequences would need to be at least $K \times L$ symbols long when a frequency reuse factor of 1 is adopted. Pilot sequences with $K \times L$ symbols are unfeasible in practical multi-cell deployments due to short channel coherence intervals. This re-use of pilot sequences causes a phenomenon known in the literature as pilot contamination, which has been considered as a major impairment in the performance of massive MIMO systems [4, 42, 69]. The effect of this phenomenon on the network is the introduction of a finite SIR, which prevents the system throughput to grow with the number of BS antennas (known as the saturation effect) [4].

We also assume, in this chapter, a wireless radio channel model that takes into account both large and small scale fading effects. Large-scale fading represents shadowing and attenuation or path-loss over a large area, while small-scale fading refers to the abrupt changes in signal amplitude and phase as a result of very small changes (in the order of $\lambda/2$) in the spatial separation between the user and the BS [103]. We consider that multipath propagation is the main effect behind small scale fading. Multipath causes multiple copies of a signal to travel over different paths with different propagation delays. These copies are received at the receiver with different phase angles and strengths [154].

The wireless channels are considered static, *i.e.*, constant, during the channel coherence interval and independent across all the M antennas, K users and coherence intervals. We begin our derivation by defining $g_{ilkm}^{(p)}$ as the complex gain of the pth path belonging to the channel connecting the k-th user in the l-th cell to the m-th antenna of the BS in the i-th cell. Additionally, the complex gain $g_{ilkm}^{(p)}$ can be expressed as a complex fast-fading factor (small-scale fading coefficient) times an amplitude factor that accounts for path-loss and log-normal shadow fading (large-scale fading coefficient),

$$g_{ilkm}^{(p)} = h_{ilkm}^{(p)}\sqrt{\beta_{ilk}}, \tag{6.1}$$

where $i = 1, \cdots L$; $l = 1, \cdots L$; $k = 1, \cdots, K$; $m = 1, \cdots, M$; $p = 0, \cdots, \Psi_P - 1$ and the fast-fading coefficients, $h_{ilkm}^{(p)}$, are assumed to follow a circularly-symmetric complex normal distribution with $\mathcal{CN}(0,1)$. The amplitude factor, $\sqrt{\beta_{ilk}}$, is assumed constant with respect to the index of the base station antenna since the path-loss and shadow fading change slowly over space [4], [154, 155]. Additionally, we assume that the set of large-scale fading coefficients, represented by $\{\beta_{ilk}\}$, is deterministic during the channel estimation. This assumption is in accordance with other works in the literature [4, 102, 131, 143].

The multipath channels connecting the k-th user in the l-th cell to the m-th antenna at the BS in the i-th cell are modeled as $1\times\Psi_P$ vectors, $\mathbf{g}_{ilkm} = \left[g_{ilkm}^{(0)}, g_{ilkm}^{(1)}, ..., g_{ilkm}^{(\Psi_P-1)}\right]$, where Ψ_P denotes the number of paths of the channel. Additionally, we assume that the channels have at most Ψ_P paths. The $M \times KP$ channel matrix including all the $\mathbf{g}_{ilkm}$ vectors is represented by

$$\mathbf{G}_{il} = \begin{bmatrix} \mathbf{g}_{il11} & \mathbf{g}_{il21} & \cdots & \mathbf{g}_{ilK1} \\ \mathbf{g}_{il12} & \mathbf{g}_{il22} & \cdots & \mathbf{g}_{ilK2} \\ \mathbf{g}_{il13} & \mathbf{g}_{il23} & \cdots & \mathbf{g}_{ilK3} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{g}_{il1M} & \mathbf{g}_{il2M} & \cdots & \mathbf{g}_{ilKM} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{il1} & \mathbf{G}_{il2} & \cdots & \mathbf{G}_{ilK} \end{bmatrix}. \tag{6.2}$$

This matrix contains all the channels for all K users in the l-th cell to all the M antennas at the BS in the i-th cell. The columns in $\mathbf{G}_{il}$, defined by $\mathbf{G}_{ilk}$, represent the channels from the k-th user in the l-th cell to all the M antennas at the BS in the i-th cell. The BS at the i-th cell requires knowledge of the channels of the users within its coverage area, denoted by $\mathbf{G}_{iik} : \forall k$, in order to perform detection and precoding operations.

6.2.1 UL Training

In order for the BS at the i-th cell to estimate all the M multipath channels from a given user to itself, each user is required to transmit an UL training sequence. We suppose the existence of pilot sequence reuse, which is a common scenario in massive MIMO. Additionally, we assume a pilot reuse factor of one and the existence of time-synchronization among all the L cells. In this scenario, users from all the L cells transmit the same pilot sequences at the same time and frequency resources. All these assumptions constitute the worst possible case for a massive MIMO system [4]. The k-th user has their pilot sequence represented by $\mathbf{s}_k = [s_k(0), s_k(1), \cdots, s_k(N-1)]^T$, where N denotes the length of the pilot sequences.

As we consider time-varying frequency-selective fading channels, each one of the K UL training sequences has to be convolved with all the M channels connecting a user to a specific BS. Therefore, with the purpose of expressing these convolutions in matrix form, we define the $N \times \Psi_P$ matrix, $\mathbf{S}_k$, with circularly-shifted versions of the pilot sequence $\mathbf{s}_k$ as shown below

$$\mathbf{S}_k = \begin{bmatrix} s_k(0) & s_k(N-1) & \cdots & s_k(N-\Psi_P+1) \\ s_k(1) & s_k(0) & \cdots & s_k(N-\Psi_P+2) \\ s_k(2) & s_k(1) & \cdots & s_k(N-\Psi_P+3) \\ \vdots & \vdots & \ddots & \vdots \\ s_k(N-1) & s_k(N-2) & \cdots & s_k(N-\Psi_P+N) \end{bmatrix}. \tag{6.3}$$

This is a circulant matrix with orthogonality property $\mathbf{S}_k^H \mathbf{S}_k = N\mathbf{I}_p$. In order to obtain circular matrices as shown in (6.3), users have to transmit their pilot sequences, $\mathbf{s}_k$, twice, rendering them UL training sequences with length $2N$, as shown in Fig. 6.2. This is mainly due to the orthogonality property exhibited by (6.3). The pilot sequences are formed through the application of cyclic shifts to ZC root sequences with prime length N. It is worth mentioning that cyclically shifted versions of a ZC sequence are orthogonal to each other [124]. ZC sequences are used in this work due to their properties of constant amplitude and zero autocorrelation, however, any other type of sequence could have been adopted as long as it exhibits the properties mentioned earlier. In order for all the users in a given cell to have valid pilots, the inequality $N > KP$ must hold. For the case when $N \geq KPL$ pilot contamination vanishes as each one of the users in the L cells will have sequences that are orthogonal to all other pilot sequences in all other cells, however, this case could be impractical as the pilot sequence could get very long depending on the number of cells, L. It is also important to mention that the number of pilot symbols, N, must be the closest prime number to KP. In case $N \leq KP$, it is not possible to form orthogonal pilot sequences out of the same ZC root sequence for all users. Additionally, the set of pilot sequences of the K users can be described by a $N \times KP$ matrix $\mathbf{S}$ of the form $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_K]$.

The received UL training sequences at the BS in the i-th cell can be represented as a $M \times N$ matrix defined as

$$\mathbf{Y}_i = \sqrt{\rho} \sum_{l=1}^{L} \mathbf{G}_{il} \mathbf{S}^H + \mathbf{N}_i, \tag{6.4}$$

where ρ is regarded as the Tx-SNR and $\mathbf{N}_i$ represents a $M \times N$ receiver noise matrix, independent over the antennas and with independent and identically distributed elements following a circularly-symmetric complex normal distribution, $\mathcal{CN}(0, 1)$.

Equation (6.4) can be written as shown in (6.5), which clearly highlights the coherent inter-cell interference caused by users transmitting the same pilot sequences within other BSs.

$$\mathbf{Y}_i = \underbrace{\sqrt{\rho}\mathbf{G}_{ii}\mathbf{S}^H}_{\text{Desired pilot signals}} + \underbrace{\sqrt{\rho} \sum_{l=1, l \neq i}^{L} \mathbf{G}_{il}\mathbf{S}^H}_{\text{Undesired pilot signals}} + \underbrace{\mathbf{N}_i}_{\text{Noise}}. \tag{6.5}$$

6.2.2 LS Channel Estimator

Each BS de-spreads the received signal, $\mathbf{Y}_i$, through a correlation operation with each one of the K pilot sequences. This operation is equivalent to right multiplying $\mathbf{Y}_i$ by $\mathbf{S}_k$. As shown in the last section, $\mathbf{S}_k$ represents the k-th pilot matrix of the overall matrix, $\mathbf{S}$. Therefore, a sufficient statistic for estimating the channel $\mathbf{G}_{ilk}$ at the BS in the i-th cell can be defined by

$$\begin{aligned} \mathbf{Z}_{ik} &= \frac{1}{\sqrt{\rho}N}\mathbf{Y}_i\mathbf{S}_k = \sum_{l=1}^{L} \mathbf{G}_{ilk} + \frac{\mathbf{N}_i\mathbf{S}_k}{\sqrt{\rho}N} \\ &= \underbrace{\mathbf{G}_{iik}}_{\text{Desired channel}} + \underbrace{\sum_{l=1, l \neq i}^{L} \mathbf{G}_{ilk}}_{\text{Inter-cell interference}} + \underbrace{\mathbf{W}_{ik}}_{\text{Noise}}, \end{aligned} \tag{6.6}$$

where $\mathbf{Z}_{ik}$ is a $M \times \Psi_P$ matrix and $\mathbf{W}_{ik} = \frac{\mathbf{N}_i\mathbf{S}_k}{\sqrt{\rho}N}$ is a $M \times \Psi_P$ noise matrix. Each one of $\mathbf{Z}_{ik}$'s columns follows a $\mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$ distribution where

$$\zeta_{ik} = \sum_{l=1}^{L} \beta_{ilk} + \frac{1}{\rho N}. \tag{6.7}$$

No information is lost during the de-sperading of the received signal once the multiplication of $\mathbf{Y}_i$ by any matrix in the orthogonal complement of $\mathbf{S}_k$ would also result in another matrix that is statistically independent of both $\sum_{l=1}^{L} \mathbf{G}_{ilk}$ and $\mathbf{W}_{ik}$ [41].

As can be noticed, ζ_{ik} is the term representing the combined interference, *i.e.*, the summation of all large-scale fading coefficients, plus noise power from the k-th users with the same pilot sequence to the BS at the i-th cell. As it appears by analyzing (6.6), the interfering channels leak directly into the desired channel estimate. The quality of the estimation is therefore limited by the SIR at the BS, which in turn limits the ability to design effective combining and precoding schemes.

Additionally, the term corresponding to noise, $\mathbf{W}_{ik}$, in (6.6) has each one of its columns following a $\mathcal{CN}(\mathbf{0}_M, \frac{1}{\rho N}\mathbf{I}_M)$ distribution.

Conventional channel estimation relies on correlating the received signal with the known pilot sequence of the kth user. This kind of estimator is referred to in the literature as the LS estimator. Hence, using (6.5), the LS estimator for the channel of the kth user within the ith cell to the ith BS is given by [126]

$$\hat{\mathbf{G}}_{iik}^{\mathrm{LS}} = \frac{1}{\sqrt{\rho}} \frac{\mathbf{Y}_i \mathbf{S}_k}{\mathbf{S}_k^H \mathbf{S}_k} = \frac{1}{\sqrt{\rho} N} \mathbf{Y}_i \mathbf{S}_k = \mathbf{Z}_{ik}. \tag{6.8}$$

The LS channel estimator has its MSE per antenna and per channel path defined by

$$\eta_{ik}^{\mathrm{LS}} = \frac{1}{M \Psi_P} \operatorname{Tr}\left\{ \mathbb{E}\left[\left(\hat{\mathbf{G}}_{iik}^{\mathrm{LS}} - \mathbf{G}_{iik} \right) \left(\hat{\mathbf{G}}_{iik}^{\mathrm{LS}} - \mathbf{G}_{iik} \right)^H \right] \right\} = \zeta_{ik} - \beta_{iik}. \tag{6.9}$$

When compared to the MMSE channel estimator, the LS estimator presents higher MSE, however, on the other hand, it does not require previous knowledge of the set of large-scale fading coefficients, $\beta_{ilk}, \forall l$. Although presenting higher MSE than the MMSE estimator, the low complexity and simplicity of the LS estimator are important aspects that can justify its adoption in real-time embedded systems. As can be noticed, the MSE, η_{ik}^{LS}, decreases with increasing ρ and decreasing β_{ilk}, $\forall i \neq l$ (*i.e.*, smaller interference level).

Remark 22. *As a result of pilot contamination, as $\rho \to \infty$, $\eta_{ik}^{LS} \to \sum_{l=1, l \neq i}^{L} \beta_{ilk}$.*

6.2.3 MMSE Channel Estimator

Statistical Bayesian MMSE channel estimation supposes previous knowledge of interference, *i.e.*, large-scale fading coefficients and noise statistics. In general, MMSE channel estimation provides simpler analytical results for lower-bound error performance, and a large number of works in massive MIMO literature employs it in order to gain knowledge of the channels [69, 130]. Those works suppose perfect knowledge of the interference and noise statistics, which might not be reasonable in practice [103]. Therefore, based on the assumption of perfect knowledge of the set of large-scale fading coefficients, $\{\beta_{ilk}\}$, and noise statistics, the ideal MMSE estimator is defined by [126]

$$\hat{\mathbf{G}}_{iik}^{\mathrm{MMSE}} = \frac{\beta_{iik}}{\zeta_{ik}} \mathbf{Z}_{ik}, \tag{6.10}$$

where each one of its columns, $\hat{\mathbf{g}}_{ilk}^{\text{MMSE }(p)}$, follows the $\mathcal{CN}(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M)$ distribution.

The ideal MMSE channel estimator has its MSE per antenna and per channel path defined by [126]

$$\eta_{ik}^{\text{MMSE}} = \frac{1}{M\Psi_P}\operatorname{Tr}\left\{\mathbb{E}\left[\left(\hat{\mathbf{G}}_{iik}^{\text{MMSE}} - \mathbf{G}_{iik}\right)\left(\hat{\mathbf{G}}_{iik}^{\text{MMSE}} - \mathbf{G}_{iik}\right)^H\right]\right\} = \beta_{iik}\left(1 - \frac{\beta_{iik}}{\zeta_{ik}}\right). \tag{6.11}$$

As can be noticed, the MSE, η_{ik}^{MMSE}, decreases with increasing ρ, decreasing β_{iik}, or decreasing β_{ilk}, $\forall i \neq l$ (*i.e.*, smaller interference level).

Remark 23. *As a result of pilot contamination, as $\rho \rightarrow \infty$, $\eta_{ik}^{MMSE} \rightarrow \beta_{iik}\left(1 - \frac{\beta_{iik}}{\sum_{l=1}^{L}\beta_{ilk}}\right)$.*

From (6.10) it is possible to see that

$$\hat{\mathbf{G}}_{ilk}^{\text{MMSE}} = \frac{\beta_{ilk}}{\beta_{iik}}\hat{\mathbf{G}}_{iik}^{\text{MMSE}}. \tag{6.12}$$

The channel estimate vector, $\hat{\mathbf{g}}_{ilk}^{\text{MMSE }(p)}$, can be expressed as

$$\hat{\mathbf{g}}_{ilk}^{\text{MMSE }(p)} = \tilde{\mathbf{g}}_{ilk}^{\text{MMSE }(p)} + \mathbf{g}_{ilk}^{(p)}, \forall p, \tag{6.13}$$

where $\tilde{\mathbf{g}}_{ilk}^{\text{MMSE }(p)}$ is the estimation error vector. Therefore, we have

$$\hat{\mathbf{g}}_{ilk}^{\text{MMSE }(p)} \sim \mathcal{CN}\left(\mathbf{0}_M, \frac{\beta_{ilk}^2}{\zeta_{ik}}\mathbf{I}_M\right), \forall p, \tag{6.14}$$

and

$$\tilde{\mathbf{g}}_{ilk}^{\text{MMSE }(p)} \sim \mathcal{CN}\left(\mathbf{0}_M, \beta_{ilk}\left(1 - \frac{\beta_{ilk}}{\zeta_{ik}}\right)\mathbf{I}_M\right), \forall p. \tag{6.15}$$

For MMSE channel estimation, the channel estimate vector, $\hat{\mathbf{g}}_{ilk}^{\text{MMSE }(p)}$, is known to be independent of the channel estimation error vector, $\tilde{\mathbf{g}}_{ilk}^{\text{MMSE }(p)}$ [126].

6.3 Proposed Channel Estimator

The process of obtaining the inter-cell large-scale fading coefficients is unjustified given the unreasonable overburden necessary to estimate them [135]. For example, assuming the deployment of L cells where each cell has K users, one given BS would have to estimate $K(L-1)$ inter-cell large-scale fading coefficients. Through a careful analysis of the ideal MMSE channel estimator, it is possible to notice that instead of the individual large-scale fading coefficients, β_{ilk}, as assumed in many previous works [4,69,76,102,129–131], it is only needed to estimate the term ζ_{ik}, which represents the sum of all large-scale fading coefficients plus noise statistic. This straightforward observation produces a simpler and efficient answer to the problem of acquiring the inter-cell large-scale coefficients. Therefore, the approach we put forward in this chapter consists of the estimation of ζ_{ik} and its subsequent use in the ideal MMSE channel estimator (6.10), resulting in the channel estimator derived next.

A possible MVUE of ζ_{ik}, given the observed signal $\mathbf{Z}_{ik}$ is defined by [126]

$$\hat{\zeta_{ik}} = \frac{\|\mathbf{Z}_{ik}\|_F^2}{M\Psi_P}, \tag{6.16}$$

where $\|\mathbf{Z}_{ik}\|_F^2 = \mathrm{Tr}\left(\mathbf{Z}_{ik}^H \mathbf{Z}_{ik}\right)$ is the squared Frobenius norm of $\mathbf{Z}_{ik}$. It is important to note that the Trace operator, Tr(.), eliminates, setting to 0, all the inner products between estimated channels of the k-th user to the M antennas at the BS in the i-th cell while summing all the elements on the main diagonal. Other possible MVUE's would be the ones that take into account one or more (less than Ψ_P) elements in the main diagonal of $\mathbf{Z}_{ik}^H \mathbf{Z}_{ik}$.

Equation (6.16) arises from the fact that $\mathbb{E}\left\{\mathbf{Z}_{ik}^H \mathbf{Z}_{ik}\right\} = M\zeta_{ik}\mathbf{I}_{\Psi_P}$, that the unbiased estimator has $\mathbb{E}\{\hat{\zeta_{ik}}\} = \zeta_{ik}$ and also from the observation that as Ψ_P increases, the term $\|\mathbf{Z}_{ik}\|_F^2/\Psi_P$ tends to $M\zeta_{ik}$, therefore resulting in a better estimator in the sense that it approaches ζ_{ik} faster than the other possible MVUE's once it is the average over all Ψ_P inner products in the main diagonal of $\mathbf{Z}_{ik}^H \mathbf{Z}_{ik}$.

It is possible to write $\zeta_{ik} = \hat{\zeta_{ik}} + e_{ik}$ where e_{ik} is the estimation error. The estimator $\hat{\zeta_{ik}}$ has Gamma distribution with shape (k) equal to $M\Psi_P$, scale (θ) equal to $\zeta_{ik}/M\Psi_P$, $\mathbb{E}\{\hat{\zeta_{ik}}\} = \zeta_{ik}$ and $\mathrm{var}\{\hat{\zeta_{ik}}\} = \zeta_{ik}^2/M\Psi_P$. Under the MVUE framework, the parameter to be estimated is considered to be deterministic [126], and thus e_{ik} has zero mean and variance given by $\zeta_{ik}^2/M\Psi_P$, which is the same variance of $\hat{\zeta_{ik}}$.

Finally, swapping ζ_{ik} with $\hat{\zeta_{ik}}$ in (6.10) yields the proposed channel estimator, which is given by

$$\hat{\mathbf{G}}_{iik}^{\mathrm{prop}} = M\Psi_P \frac{\beta_{iik}}{\|\mathbf{Z}_{ik}\|_F^2} \mathbf{Z}_{ik}. \tag{6.17}$$

As will be shown next, the proposed channel estimator asymptotically approaches the ideal MMSE channel estimator with regard to both M and Ψ_P. The estimated channel matrix, $\hat{\mathbf{G}}_{iik}^{\mathrm{prop}}$, can be expressed as

$$\hat{\mathbf{G}}_{iik}^{\mathrm{prop}} = [\hat{\mathbf{g}}_{iik}^{\mathrm{prop}\ (0)}, \hat{\mathbf{g}}_{iik}^{\mathrm{prop}\ (1)}, \ \ldots, \ \hat{\mathbf{g}}_{iik}^{\mathrm{prop}\ (\Psi_P - 1)}], \tag{6.18}$$

where $\hat{\mathbf{g}}_{iik}^{\mathrm{prop}\ (p)}$ $\forall p$ are $M \times 1$ vectors containing the channel coefficients of the pth path where each one of these vectors has $\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop }(p)}] = \mathbf{0}_M$ and covariance matrix given by

$$\begin{aligned}\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop }(p)}(\hat{\mathbf{g}}_{iik}^{\text{prop }(p)})^H] &= \left(\frac{M\Psi_P}{M\Psi_P - 1}\frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\ &= \left(\frac{\beta_{iik}^2}{\zeta_{ik}} + \frac{1}{(M\Psi_P - 1)}\frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\ &= \left(\frac{\beta_{iik}^2}{\zeta_{ik}} + \epsilon_{ik}\right)\mathbf{I}_M, \forall p,\end{aligned} \tag{6.19}$$

where $\epsilon_{ik} = \frac{1}{(M\Psi_P-1)}\frac{\beta_{iik}^2}{\zeta_{ik}}$. The mean of $\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop }(p)}], \forall p$ can be found by using the symmetry property of the distribution of $z_{ikm}^{(p)} \sim \mathcal{CN}(0, \zeta_{ik})$, we conclude that $\mathbb{E}\left[\hat{g}_{iikm}^{\text{prop }(p)}\right] = 0$, once $\mathbb{E}\left[\frac{z_{ikm}^{(p)}}{\|\mathbf{z}_{ik}\|_F^2}\right] = \mathbb{E}\left[\frac{-z_{ikm}^{(p)}}{\|\mathbf{z}_{ik}\|_F^2}\right], \forall m, p.$

As can be seen by analyzing (6.19), as $M\Psi_P \to \infty$, $\text{Var}[\hat{\mathbf{g}}_{iik}^{\text{prop }(p)}] \to \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M$, which is the covariance matrix of the MMSE estimator. The proof for the covariance matrix of the proposed estimator is presented in Appendix H.

For the case when $\Psi_P = 1$ (*i.e.*, the channel can be considered as being flat throughput the transmission bandwidth) it is easy to notice that (6.17) reduces to (4.16) presented in chapter 4. An approximation to the MSE per antenna and per channel path of the proposed channel estimator is defined by

$$\eta_{ik}^{\text{prop}} = \frac{1}{M\Psi_P}\text{Tr}\left\{\mathbb{E}\left[\left(\hat{\mathbf{G}}_{iik}^{\text{prop}} - \mathbf{G}_{iik}\right)\left(\hat{\mathbf{G}}_{iik}^{\text{prop}} - \mathbf{G}_{iik}\right)^H\right]\right\} \approx \beta_{iik}\left[1 - \frac{(M\Psi_P - 2)\beta_{iik}}{(M\Psi_P - 1)\zeta_{ik}}\right]. \tag{6.20}$$

The approximate MSE for the proposed estimator, η_{ik}^{prop}, decreases with increasing ρ, increasing $M\Psi_P$, decreasing β_{iik}, or decreasing β_{ilk}, which means smaller interference level from other cells, *i.e.*, smaller pilot contamination.

Remark 24. *As a result of pilot contamination, as $\rho \to \infty$ and $M\Psi_P \to \infty$, $\eta_{ik}^{prop} \to \beta_{iik}\left(1 - \frac{\beta_{iik}}{\sum_{l=1}^{L}\beta_{ilk}}\right)$.*

Remark 24 clearly shows that the MSE of the proposed channel estimator tends to the MSE of the ideal MMSE channel estimator when both ρ and $M\Psi_P \to \infty$. The proof for the approximation of the MSE is given in Appendix I.

Remark 25. *The mean squared Euclidean distance per antenna and per channel between $\hat{\boldsymbol{G}}_{iik}^{prop}$ and $\hat{\boldsymbol{G}}_{iik}^{MMSE}$ is given by*

$$\frac{1}{M\Psi_P}\text{Tr}\left\{\mathbb{E}\left[\left(\hat{\mathbf{G}}_{iik}^{\text{prop}} - \hat{\mathbf{G}}_{iik}^{\text{MMSE}}\right)\left(\hat{\mathbf{G}}_{iik}^{\text{prop}} - \hat{\mathbf{G}}_{iik}^{\text{MMSE}}\right)^H\right]\right\} = \frac{1}{(M\Psi_P - 1)}\frac{\beta_{iik}^2}{\zeta_{ik}} = \epsilon_{ik}. \tag{6.21}$$

The proof of Remark 25 is presented in Appendix J. After analyzing equations (6.7) and (6.21) it is easily noticeable that the mean squared Euclidean distance in (6.21) decreases as ρ decreases, $M\Psi_P$ increases, β_{iik} decreases, and $\beta_{ilk},\ i \neq l$ increases.

Remark 26. *The channel estimation error, $\tilde{\boldsymbol{G}}_{iik}^{prop} = \hat{\boldsymbol{G}}_{iik}^{prop} - \boldsymbol{G}_{iik}$, is correlated with the channel estimate and uncorrelated with the pilot signal, $\boldsymbol{Z}_{ik}$,*

$$\frac{1}{M}\mathbb{E}\left[(\tilde{\mathbf{G}}_{iik}^{\text{prop}})^H \hat{\mathbf{G}}_{iik}^{\text{prop}}\right] = \epsilon_{ik}\mathbf{I}_{\Psi_P}, \tag{6.22}$$

$$\frac{1}{M}\mathbb{E}\left[(\tilde{\mathbf{G}}_{iik}^{\text{prop}})^H \mathbf{Z}_{ik}\right] = \mathbf{0}_{\Psi_P \times \Psi_P}, \tag{6.23}$$

The proof of (6.22) follows the same rationale used to find the third term of (I.1) in Appendix I. It is clear that (6.22), like (6.21), decreases with increasing $M\Psi_P$, decreasing ρ, increasing $\beta_{ilk}, i \neq l$, and decreasing β_{iik}. The estimation error vector, $\tilde{\mathbf{g}}_{iik}^{\text{prop}\ (p)}$, has the following mean vector and covariance matrix,

$$\mathbb{E}[\tilde{\mathbf{g}}_{iik}^{\text{prop}\ (p)}] = \mathbf{0}_M, \forall p, \tag{6.24}$$

$$\begin{aligned}
\text{Cov}\left(\tilde{\mathbf{g}}_{iik}^{\text{prop}\ (p)}\right) &= \mathbb{E}[\tilde{\mathbf{g}}_{iik}^{\text{prop}\ (p)}(\tilde{\mathbf{g}}_{iik}^{\text{prop}\ (p)})^H] \\
&\approx \left(\beta_{iik} + \epsilon_{ik} - \frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\
&= \beta_{iik}\left[1 - \frac{(M\Psi_P - 2)\beta_{iik}}{(M\Psi_P - 1)\zeta_{ik}}\right] \\
&= \left[\beta_{iik}\left(1 - \frac{\beta_{iik}}{\zeta_{ik}}\right) + \epsilon_{ik}\right]\mathbf{I}_M, \forall p.
\end{aligned} \tag{6.25}$$

Remark 27. *As $M\Psi_P \to \infty$ the covariance matrix of the estimation error, $\tilde{\boldsymbol{g}}_{iik}^{prop\ (p)}$, tends to $\mathbb{E}\left[\tilde{\boldsymbol{g}}_{iik}^{MMSE\ (p)}\left(\tilde{\boldsymbol{g}}_{iik}^{MMSE\ (p)}\right)^H\right] = \left[\beta_{iik}\left(1 - \frac{\beta_{iik}}{\zeta_{ik}}\right)\right]\boldsymbol{I}_M, \forall p.$*

6.4 Simulation Results

A thorough analysis of the performance of the proposed channel estimator is carried out in this section. We use the MMSE and LS channel estimators for some of the comparisons with the proposed channel estimator. A usual multi-cell deployment as depicted in Fig. 6.1 with the parameters described in Table 6.1 is adopted in this work.

We consider two different cases for setting the large-scale fading coefficients, $\{\beta_{ilk}\}$, one considering constant values and other considering random values. In the case where the coefficients are considered constant, β_{iik} is set to 1 and $\beta_{ilk}, \forall\ l \neq i$ is set to

Table 6.1 – Simulation Parameters.

Parameter	Description	Value
L	Total number of cells	7
K	Number of users per cell	10
Ψ_P	Number of multipath coefficients per channel	20
N	Pilot length in symbols	223
-	Frequency and pilot reuse factors	1

a. In the case where the coefficient's values are randomly selected, the users belonging to each one of the L cells are uniformly distributed within the area comprised by two circles with radius $d_0 = 100$ m and $d_1 = 1000$ m respectively. The set of large-scale fading coefficients $\{\beta_{ilk}\}$ is independently generated by $\beta_{ilk} = \psi / \left(\frac{d_{ilk}}{d_0}\right)^v$, where $v = 3.8$, $10 \log_{10}(\psi) \sim \mathcal{N}(0, \sigma^2_{\text{shadow,dB}})$ with $\sigma_{\text{shadow,dB}} = 8$, and d_{ilk} is the distance of the k-th user in l-th cell to BS in the i-th cell.

Fig. 6.3 (a) depicts the channel estimation MSE versus the transmit signal to noise ratio, ρ, for $M = 30$ and $a = 0.05$. As is evident from the figure, both simulation and analytical MSE results match for all the compared estimators. The MSE of all channel estimators decreases with the increase of the transmit signal to noise ratio. It can also be easily noticed that all the three channel estimators present MSE floors due to pilot contamination. This confirmation is in accordance with Remarks 22, 23 and 24. The MSE of the proposed channel estimator gets quite close to the MSE of the ideal MMSE channel

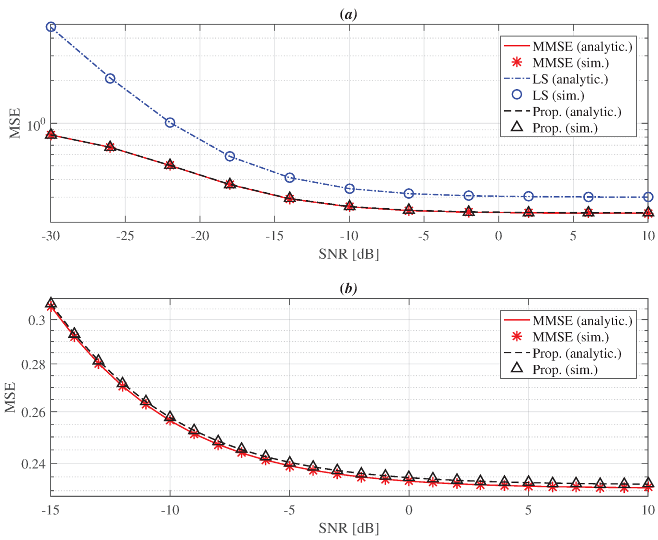

Figure 6.3 – MSE of channel estimation versus UL pilot power: (a) for $M = 30$ and $a = 0.05$, (b) shows that gap between the MMSE and the proposed channel estimators increases, tending to a constant value as ρ increases.

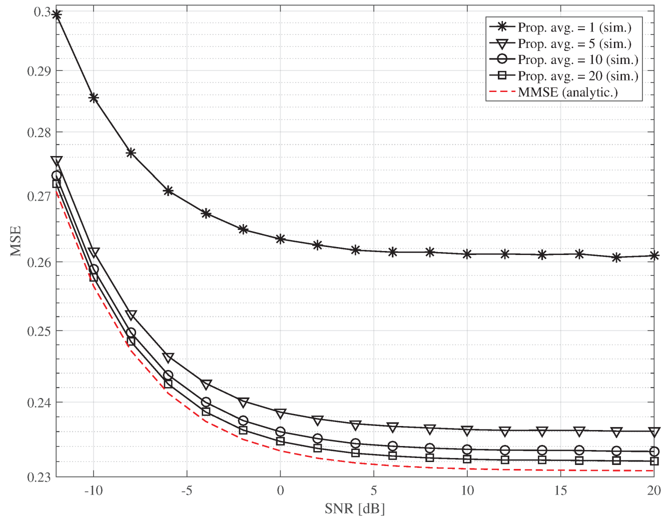

Figure 6.4 – Performance of the proposed estimator when varying the number of averaged elements.

estimator at low transmit signal to noise ratio values while the difference between the LS and both the ideal MMSE and proposed channel estimators increases. Fig. 6.3 (b) shows that the difference between the ideal MMSE channel estimator and the proposed one increases, tending to a constant value as the transmit signal to noise ratio increases. This confirmation is in accordance with Remark 25.

The impact caused by the number of averaged elements on the performance of the proposed channel estimator is depicted in Fig. 6.4. In this simulation, the number of elements on the main diagonal of $\mathbf{Z}_{ik}^{H}\mathbf{Z}_{ik}$ that are averaged to calculate the proposed channel estimator is varied. The figure clearly indicates that the performance of the proposed channel estimator is directly affected by the number of elements taken into account for the calculation of the estimator. Consequently, the greater the number of averaged ele-

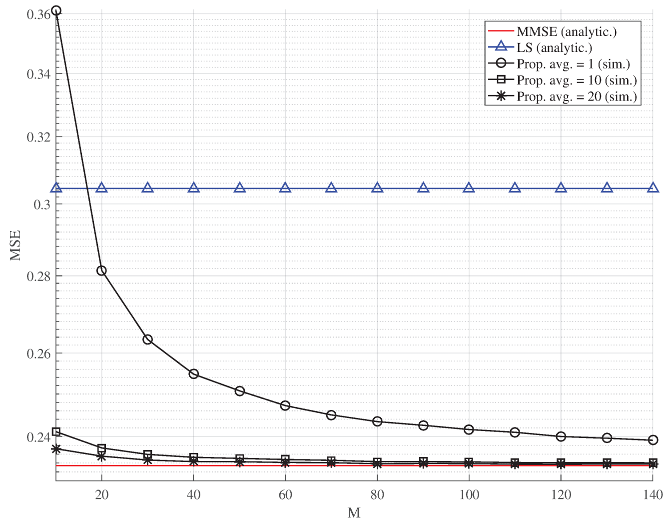

Figure 6.5 – MSE of channel estimation versus number of antennas, M.

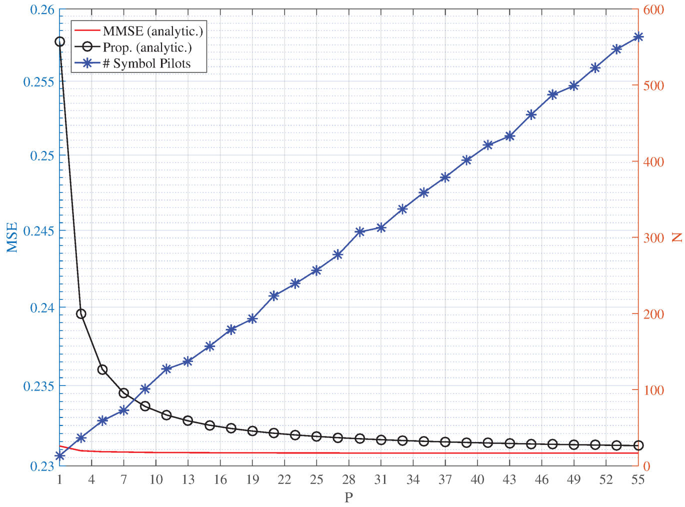

Figure 6.6 – MSE of channel estimation versus number of channel paths versus pilot length.

ments, the smaller the MSE difference between the proposed and the ideal MMSE channel estimators. Additionally, the number of elements of $\mathbf{Z}_{ik}^H\mathbf{Z}_{ik}$ considered in the average can also be thought of as the number of paths, Ψ_P, a time-varying frequency-selective fading channel has. The figure also indicates that the performance of the proposed channel estimator improves with Ψ_P. Therefore, this result demonstrates that the proposed estimator not only exploits spatial diversity but also takes advantage of the time diversity provided by the delay spread of the channels, since its performance improves with Ψ_P.

Fig. 6.5 shows the impact of varying the number of antennas, M, installed at one BS on the channel estimation MSE. For this result, we considered $a = 0.05$ and Tx-SNR, $\rho = 0$ dB. The MSE of the proposed channel estimator asymptotically approaches that of the ideal MMSE channel estimator as the number of antennas, M, increases. On the other hand, the channel estimation MSE of the LS channel estimator

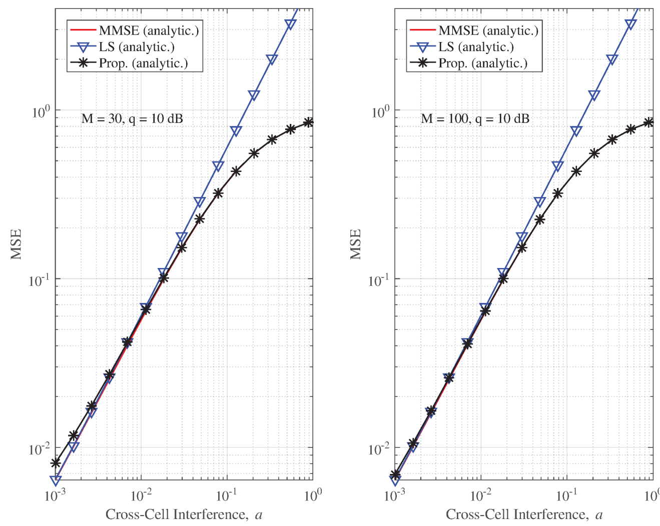

Figure 6.7 – MSE of channel estimation versus inter-cell interference, a.

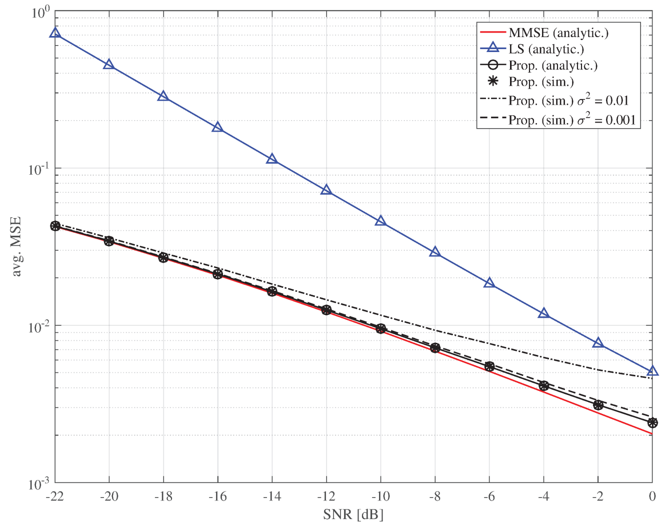

Figure 6.8 – Averaged MSE of channel estimation for random$\{\beta_{ilk}\}$.

stays fixed at a constant value for all values of M. In this figure, we also show the impact of the number of averaged elements of $\mathbf{Z}_{ik}^H\mathbf{Z}_{ik}$ on the performance of the proposed channel estimator. From the results, it is observable that the proposed channel estimator has its MSE performance improved as the number of elements from the main diagonal of $\mathbf{Z}_{ik}^H\mathbf{Z}_{ik}$ taken into account in the average increases, *i.e.*, as time diversity, represented here by Ψ_P, increases.

Fig. 6.6 depicts the comparison of channel estimation MSE versus the number of channel paths, Ψ_P, versus the number of pilot symbols, N, for both the ideal MMSE and proposed channel estimators. For this result, we considered the number of BS antennas, $M = 30$ and Tx-SNR, $\rho = 20$ dB. The figure shows that as the number of channel paths, Ψ_P, increases, the MSE of the proposed channel estimator asymptotically approaches that of the ideal MMSE channel estimator. It is also shown that as Ψ_P increases, the length of

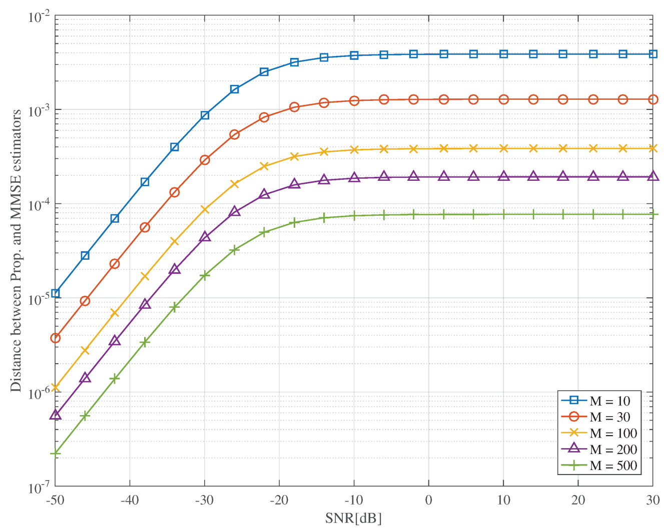

Figure 6.9 – Distance between proposed and MMSE estimators (Remark 25).

the pilot sequence, N, has to be also increased in order to provide users with orthogonal sequences. Without orthogonal sequences, the BS is not able to estimate the channels from its users to itself properly.

Fig. 6.7 shows the impact of different levels of inter-cell interference, here denoted by a, on the performance of channel estimation MSE for the three estimators considered in this work. For this result, we considered two different number of BS antennas, $M = 30$ and 100 and Tx-SNR, $\rho = 10$ dB. When compared with the proposed channel estimator, the LS estimator presents a slightly better MSE result for low levels of inter-cell interference and a small number of BS antennas, M, *e.g.*, $M = 30$. However, as M increases, this difference disappears. It can be clearly noticed in the plot with $M = 100$. It is clear from both plots that the proposed channel estimator substantially surpasses the LS estimator and approaches the performance of the ideal MMSE estimator as the inter-cell interference level increases. This observation is in accordance with Remark 25.

Fig. 6.8 compares the performance of channel estimation MSE for the case when the set of large-scale fading coefficients, $\{\beta_{ilk}\}$, is randomly selected. For this result, we considered the number of BS antennas, $M = 30$. The MSE results shown in the figure are the result of averaging the MSE values over 10000 realizations of the set $\{\beta_{ilk}\}$. As can be noticed, the approximated analytical MSE expression matches the simulated MSE results. Furthermore, we investigate the sensitivity of the proposed channel estimator against inaccuracy in the estimation of the intra-cell large-scale fading coefficients, β_{iik}, when we use the estimate $\beta_{iik} = \beta_{iik}(1 + \mathcal{N}(0, \sigma^2))$. The degradation in the performance of the proposed estimator becomes quite clear at high Tx-SNR values when $\sigma^2 = 0.01$, however, when $\sigma^2 = 0.001$ it becomes negligible. The performance of the proposed channel estimator still surpasses that of the LS estimator significantly.

In Fig. 6.9, we compare the distance between the proposed and MMSE channel estimators for different number of antennas, M, with $a = 0.05$. As Remark 25 states, the distance between them is small at low SNR, increasing with SNR until a ceiling value is reached. As can be also noticed, the ceiling value decreases as the number of antennas, M, increases. Also aligned with the interpretation of (6.21), the figure shows that increasing the pilot power does not improve the performance of the proposed estimator.

6.5 Conclusions

This chapter has proposed the use of Zadoff-Chu pilot sequences for the UL training phase of massive MU-MIMO TDD cellular systems, in such a way that we could derive a simple and practical channel estimator for the respective frequency-selective massive MIMO fading channels, assuming pilot contamination. The estimator proposed in this chapter is defined by replacing the combined interference plus noise power term, ζ_{ik}, in the ideal MMSE estimator with an MVUE estimator for that term. The resulting channel estimator has proved to be simpler and more practical than the ideal MMSE channel estimator. Additionally, we derive and evaluate an approximate analytical MSE expression for the proposed channel estimator. Through analytical analysis and extensive simulations, the resulting channel estimator has proved to be simpler and more practical than the ideal MMSE estimator. Its MSE results asymptotically approach that of the ideal MMSE estimator without any previous knowledge of the inter-cell large-scale coefficients and noise power term, ζ_{ik}.

7 Concluding Remarks

In this thesis, we presented a set of contributions that address the UL mixed services problem and issues related to channel estimation in multi-cell scenarios. The findings presented here shed light on and motivate for entirely new research lines towards a better understanding of massive MIMO systems and also their application to MTC networks. Therefore, as the discussion in this thesis attests, massive MIMO technology presents the potential to allow the coexistence of MTC and HTC devices, thanks to its inherent spatial multiplexing properties and low transmission power requirements. Additionally, we show that it is possible to estimate channels in multi-cell scenarios, which are heavily affected by the pilot contamination phenomenon, without the need to know beforehand the channel statics. Next, the main contributions of this thesis are summarized.

- In chapter 2, we presented a survey on existing related work on massive MIMO systems and identified some of the main challenges and their possible solutions. We investigated the feasibility of applying Massive MIMO as a means to address the so-called UL mixed-service communication problem, where a BS has to serve HTC devices and a great number of MTC devices. We propose the PNSCH channel, which is devised to consume the data traffic generated by MTC devices. A discussion on the possibility of applying simple and sub-optimal linear detection schemes in the UL is also presented. Additionally, based on perfect CSI we derive closed-form lower-bound expressions on the UL achievable rates for each one of the studied detectors.
- In chapter 3, we continue the discussion started in chapter 2 by extending it to the imperfect CSI case. We show that the UL mixed-service communication problem can be split into two sub-problems, namely, random access and data transmission problems. The focus of this chapter is on the data transmission phase. A channel estimator for single-cell scenarios (or systems with frequency reuse-factors greater than 1) that does not need previous information on the channel statics is devised by employing the ML method. We prove that the proposed estimator is unbiased and achieves the Crámer-Rao lower bound. Finally, we derive lower bounds on the achievable rate for each one of the studied detectors. Our results show that even for sub-optimal linear receivers, the transmitted power of the MTC devices can be reduced as the number of antennas grows without bound.
- In chapter 4, we initially study the effects of pilot contamination on the channel

estimation performance and then we propose and evaluate the performance of a simple and effective channel estimator used to mitigate the pilot contamination problem. The proposed channel estimator does not need prior knowledge of both the large-scale fading coefficients of the interfering cells and the noise power. Our results show that the performance of the proposed estimator asymptotically approaches that of the ideal MMSE estimator as the number of antennas grows without limit.

- In chapter 5, we presented a study on the distribution of the channel estimator proposed in chapter 4. We proved that its distribution can be accurately approximated by the circularly-symmetric complex normal distribution for a BS equipped with more than 10 antennas, which is a feasible value given the current massive MIMO deployments.
- In chapter 6, we propose an UL training scheme that employs ZC sequences to ease the channel estimation process in SC-FDMA systems. We study the problem of estimating channels on multi-cell systems that use SC-FDMA transmissions. Based on the proposed estimation scheme, we employ the MVUE method to estimate the large-scale fading coefficients plus noise power term and replace it into the ideal MMSE channel estimator, resulting in an efficient, more practical and simpler channel estimator. An approximate analytical MSE expression for the proposed estimator is also presented and its accuracy is attested through simulations. Our numerical results demonstrate that the proposed channel estimator performs asymptotically as well as the ideal MMSE channel estimator as the number of antennas and channel paths increases.

7.1 Future Work

Next, we present a list of possible directions for future work, related to our study and findings described in this thesis, which can yield very promising research results.

7.1.1 Random Access for Massive Machine-Type Communications

In chapters 2 and 3 we focus our attention on the data transmission phase and propose a scheme to accommodate the data transmissions of a great number of MTC devices through the use of massive MIMO technology. On the other hand, the random access phase problem poses several challenges is also of great research interest. Therefore, a possible future direction would be devising new mechanisms to grant network access to a huge number of devices (HTC and MTC) competing for resources.

7.1.2 Spatially Correlated Channels

Practical channels are not spatially uncorrelated. Firstly, due to the propagation environment, the channel vector is more probable to point in some directions than in others. Secondly, the antennas have spatially dependent antenna patterns. Both factors contribute to the fact that spatial channel correlation always appears in practice [158]. Therefore, as another future work direction, the study of the impact that spatially correlated fading (Rician and Rayleigh fading), *i.e.*, spatially correlated channels, has on channel estimation, linear detection and SE are also of great interest for the research community.

7.1.3 Detection for SC-FDMA Systems

In chapter 6 we studied the channel estimation problem in multi-cell massive MIMO systems employing SC-FDMA modulation. However, another interesting research problem related to SC-FDMA massive MIMO systems is the detection of transmitted symbols. A very interesting research topic would be the study of the feasibility of employing sub-optimal detectors in these systems and the respective SE attained by such detectors.

7.1.4 Mitigation of Pilot-Contamination

As discussed in chapters 4 and 6, in Rayleigh fading channels, pilot-contamination does not disappear even when the number of antennas deployed at the BS grows without bound. This phenomenon is caused by the fact that the coherence time is finite and consequently, the number of orthogonal pilot sequences available at the different BSs is also finite. This finite number of sequences makes it impossible for a given device to pick a sequence that is mutually orthogonal to all other sequences being used at all neighbour cells, consequently bringing about a coherent interference effect known as pilot contamination. Therefore, possible algorithms, mechanisms, etc. that can mitigate if not eliminate the pilot contamination effect are of great interest to the research community and constitute another interesting future work direction.

Bibliography

[1] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020*, White paper, February 2016.

[2] ITU-R, Recommendation ITU-R M.2083, *IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond*, September 2015.

[3] Ericsson, *Ericsson mobility report*, [Online]. Available: http://www.ericsson.com/mobility-report, June 2017.

[4] T. L. Marzetta, *Noncooperative cellular wireless with unlimited numbers of base station antennas*, IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590-3600, November 2010.

[5] Huawei Technologies CO., LTD, *5G Network Architecture - A High-Level Perspective*, White Paper, Online. [Available]: https://www.huawei.com/en/industry-insights/outlook/mobile-broadband/insights-reports/5g-network-architecture, 2016.

[6] 4G Americas White Papers, *4G Americas' Recommendations on 5G Requirements and Solutions*, Online. [Available]: http://www.4gamericas.org/en/ resources/white-papers/, October 2014.

[7] Keith Mallinson, *The path to 5G: as much evolution as revolution*, [available at] http://www.3gpp.org/news-events/3gpp-news/1774-5g_wiseharbour, May 2016.

[8] Oumer Teyeb, Gustav Wikstrom, Magnus Stattin, Thomas Cheng, Sebastian Faxer, and Hieu Do, *Evolving LTE to fit the 5G future*, Ericsson Technology Review. [available at] https://www.ericsson.com/en/ericsson-technology-review/archive/2017/evolving-lte-to-fit-the-5g-future, January 2017.

[9] Evangelos Vlachos, Aris S. Lalos, Kostas Berberidis, and Christos Tselios, *Autonomous driving in 5G: Mitigating interference in OFDM-based vehicular communications*, IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), June 2017.

[10] G. Fettweis and S. Alamouti, *5G: Personal Mobile Internet beyond What Cellular Did to Telephony*, IEEE Communications Magazine, vol. 52, no. 2, pp. 140-145, February 2014.

[11] G. Wunder et al., *5GNOW: Non-Orthogonal Asynchronous Waveforms for Future Mobile Applications*, IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, February 2014.

[12] ITU-R Report M.2370-0, *IMT traffic estimates for the years 2020 to 2030*, July 2015.

[13] J. Gubbi et al., *Internet of Things (IoT): A vision, architectural elements, and future directions*, Future Generation Computer Systems, vol. 29, no. 7, pp. 1645-1660, September 2013.

[14] D. Boswarthick, O. Elloumi, and O. Hersent, *M2M Communications: A Systems Approach*, Wiley, 2012.

[15] A. Elmangoush, H. Coskun, S. Wahle, N. Blum and T. Magedanz, *Promoting M2M Application Development for Smart City*, 29th meeting of the Wireless World Research Forum (WWRF), October 2012.

[16] J. M. Corchado, J. Bajo, D. I. Tapia, and A. Abraham, *Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare*, IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 234-240, March 2010.

[17] P. Papadimitratos, A. de La Fortelle, K. Evensen, R. Brignolo, and S. Cosenza, *Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation*, IEEE Communications Magazine, vol. 11, no. 1, pp. 84-95, November 2009.

[18] K. Moslehi and R. Kumar, *A reliability perspective of the smart grid*, IEEE Transactions on Smart Grid, vol. 1, no. 1, pp. 57-64, June 2010.

[19] Elias Kyriakides and Marios Polycarpou, *Intelligent Monitoring, Control and Security of Critical Infrastructure Systems*, Springer, 1st Ed. 2015.

[20] S. Goel, S. F. Bush, and D. E. Bakken, *IEEE Vision for Smart Grid Communications: 2030 and Beyond*, IEEE Press, E-ISBN: 978-0-7381-8460-9, May 2013.

[21] Tadilo Endeshaw Bogale and Long Bao Le, *Massive MIMO and Millimeter Wave for 5G Wireless HetNet: Potentials and Challenges*, IEEE Vehicular Technology Magazine, vol. 11, no. 1, pp. 64-75, March 2016.

[22] C. S. Pattichis, E. Kyriacou, S. Voskarides, M. S. Pattichis, and R. Istepanian, *Wireless telemedicine Systems: An Overview*, IEEE Antennas and Propagation Magazine, vol 44, no. 2, pp.143-153, April 2002.

[23] Anita Majerowicz and Susan Tracy, *Telemedicine: Bridging Gaps in Healthcare Delivery*, Journal of AHIMA, vol. 81, no. 5, pp. 52-53, May 2010.

[24] G. Araniti, C. Campolo, M. Condoluci, A. Iera, and A. Molinaro, *LTE for Vehicular Networking: A Survey*, IEEE Communications Magazine, vol. 51, no. 5, pp. 148-157, May 2013.

[25] Tamer ElBatt, Siddhartha K. Goel, Gavin Holland, Hariharan Krishnan, and Jayendra Parikh, *Cooperative collision warning using dedicated short range wireless communications*, International Workshop on Vehicular Ad Hoc Networks (VANET), pp. 1-9, September 2006.

[26] Christoph Ide, Fabian Kurtz, and Christian Wietfeld, *Cluster-Based Vehicular Data Collection for Efficient LTE Machine-Type Communication*, IEEE Vehicular Technology Conference (VTC), September 2013.

[27] Feng Jiang, Jie Chen, A. Lee Swindlehurst, and José A. López-Salcedo, *Massive MIMO for Wireless Sensing With a Coherent Multiple Access Channel*, IEEE Transactions on Signal Processing, vol. 63, no. 12, pp. 3005-3017, June 2015.

[28] Amirpasha Shirazinia, Subhrakanti Dey, Domenico Ciuonzo, and Pierluigi Salvo Rossi, *Massive MIMO for Decentralized Estimation of a Correlated Source*, IEEE Transactions on Signal Processing, vol. 64, no. 10, pp. 2499-2512, May 2016.

[29] Domenico Ciuonzo, Pierluigi Salvo Rossi, and Subhrakanti Dey, *Massive MIMO channel-aware decision fusion*, IEEE Transactions on Signal Processing, vol. 63, no. 3, pp. 604-619, February 2015.

[30] Francis da Costa, *Rethinking the Internet of Things: A Scalable Approach to Connecting Everything*, Apress, 1st Ed. 2014.

[31] G. Wunder, P. Jung, and C. Wang, *Compressive Random Access for Post-LTE Systems*, IEEE ICC - Workshop on Massive Uncoordinated Access Protocols, pp. 539-544, June 2014.

[32] B. C. Villaverde et al., *Service Discovery Protocols for Constrained Machine-to-Machine Communications*, IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 41-60, 2014.

[33] G. Wunder, M. Kasparick, and P. Jung., *Spline Waveforms and Interference Analysis for 5G Random Access with Short Message Support*, arXiv:1501.02917, Online. [Available]: http://arxiv.org/pdf/1501.02917.pdf, January 2015.

[34] Yinsheng Liu et al., *Waveform Design for 5G Networks: Analysis and Comparison*, IEEE Access, vol. 5, pp. 19282-19292, February 2017.

[35] J. G. Andrews et al., *What Will 5G Be?*, IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065-1082, June 2014.

[36] Aymen Omri, Mazen O. Hasna, and Mohammed Nafie, *Effective area spectral efficiency for wireless communication networks with interference management*, EURASIP Journal on Wireless Communications and Networking, vol. 2015, no. 1, pp. 205-217, August 2015.

[37] Theodore S. Rappaport, Yunchou Xing, George R. MacCartney, Jr., Andreas F. Molisch, Evangelos Mellios, and Jianhua Zhang, *Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks-With a Focus on Propagation Models*, IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, December 2017.

[38] Theodore S. Rappaport, *Spectrum frontiers: The new world of millimeter-wave mobile communication*, Invited Keynote Presentation, The Federal Communications Commission (FCC) Headquarters, 2016.

[39] Marja Matinmikko, et al., *Overview and comparison of recent spectrum sharing approaches in regulation and research: From opportunistic unlicensed access towards licensed shared access*, IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN), April 2014.

[40] Felipe A. P. de Figueiredo et. al., *A Spectrum Sharing Framework for Intelligent Next Generation Wireless Networks*, IEEE Access, vol. 6, pp. 60704-60735, December 2018.

[41] Thomas L. Marzetta, Erik G. Larsson, Hong Yang, Hien Quoc Ngo, *Fundamentals of Massive MIMO*, Cambridge U.K., Cambridge University Press, 1st edition, November 2016.

[42] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta., *Massive MIMO for next generation wireless systems*, IEEE Communications Magazine, vol. 52, no. 2, pp. 186-195, February 2014.

[43] Emil Bjornson, Erik G. Larsson, and Thomas L. Marzetta, *Massive MIMO: Ten Myths and One Critical Question*, IEEE Communications Magazine, vol. 54, no. 2, pp. 114-123, February 2016.

[44] Emil Bjornson, Jakob Hoydis, and Luca Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*, Foundations and Trends in Signal Processing: vol. 11, no. 3-4, pp. 154-655, DOI: 10.1561/2000000093, November 2017.

[45] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Renato R. Lopes, and Joao Paulo Miranda, *On the Application of Massive MU-MIMO in the Uplink of Machine Type Communication Systems*, International Workshop on Telecommunications (IWT), June 2015.

[46] Joao Paulo Miranda, Arman Farhang, Nicola Marchetti, Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, and Fabrício Figueiredo, *On massive MIMO and its applications to machine type communications and FBMC-based networks*, EAI Endorsed Transactions on Ubiquitous Environments, vol. 2, no. 5, July 2015.

[47] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman, Gustavo Fraidenraich, *On the Application of Massive MIMO Systems to Machine Type Communications*, IEEE Access, vol. 7, no. 12, pp. 2589-2611, December 2018.

[48] Andres Laya, Luis Alonso, and Jesus Alonso-Zarate, *Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives*, IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 4-16, First Quarter 2014.

[49] Stefania Sesia, Issam Toufik and Matthew Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, John Wiley & Sons, ISBN:9780470697160, February 2009.

[50] S.-Y. Lien, K.-C. Chen, and Y. Lin, *Toward Ubiquitous Massive Accesses in 3GPP Machine-to-machine Communications*, IEEE Communications Magazine, vol. 49, no. 4, pp. 66-74, April 2011.

[51] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A.-C. Hsu, *Overload control for machine-type-communications in LTE-advanced system*, IEEE Communications Magazine, vol. 50, no. 6, pp. 38-45, June 2012.

[52] M. Gerasimenko et al., *Energy and delay analysis of LTE-advanced RACH performance under MTC overload*, IEEE Globecom Workshops, pp. 1632-1637, December 2012.

[53] U. Phuyal, A. Koc, M.-H. Fong, and R. Vannithamby, *Controlling access overload and signaling congestion in M2M networks*, Asilomar Conference on Signals, Systems, and Computers, pp. 591-595, November 2012.

[54] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, *An Overview of Massive MIMO: Benefits and Challenges*, IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 742-758, October 2014.

[55] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, Springer, 3rd Ed. 2004.

[56] Shepard, C. et al., *Argos: Practical many-antenna base stations*, International Conference on Mobile Computing and Networking (ACM MobiCom), pp. 53-64, August 2012.

[57] Janaswamy, R., *Effect of element mutual coupling on the capacity of fixed length linear arrays*, IEEE Antennas Wireless Propagation Letters, vol. 1, no. 1, pp. 157-160, 2002.

[58] Wallace, J. W. and Jensen, M. A., *Mutual coupling in MIMO wireless systems: A rigorous network theory analysis*, IEEE Transactions on Wireless Communications, vol. 3, no. 4, pp. 1317-1325, July 2004.

[59] Lau, B. K. et al., *Impact of matching network on bandwidth of compact antenna arrays*, IEEE Transactions on Antennas Propagation, vol. 54, no. 11, pp. 3552-3238, November 2006.

[60] Sibille, A., Oestges, C. and Zanella, A., *MIMO: From Theory to Implementation*, Academic Press, 2010.

[61] Rusek, F. et al., *Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays*, IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40-60, January 2013.

[62] Moustakas, A. L. et al., *Communication through a diffusive medium: Coherence and capacity*, Science, vol. 287, pp. 287-290, January 2000.

[63] Artiga, X., Devillers, B. and Perruisseau-Carrier, J., *On the selection of radiating elements for compact indoor massive-multiple input multiple output base stations*, IET Microwaves, Antennas & Propagation, vol. 8, no. 1, pp. 1-9, January 2014.

[64] Gao, X., Tufvesson, F. and Edfors, O., *Massive MIMO channels - measurements and models*, Asilomar Conference on Signals, Systems and Computers, November 2013.

[65] Shangbin Wu, Cheng-Xiang Wang, Yang Yang, Wenjin Wang, and Xiqi Gao, *Performance comparison of massive MIMO channel models*, IEEE/CIC International Conference on Communications in China (ICCC), July 2016.

[66] Zheng, K. Ou, S. and Yin, X., *Massive MIMO Channel Models: A Survey*, International Journal of Antennas and Propagation, vol. 2014, 10 pages, June 2014.

[67] Jiang, Z. et al., *Achievable Rates of FDD Massive MIMO Systems with Spatial Channel Correlation*, IEEE Transactions on Wireless Communications, vol. 14, no. 5, pp. 2868-2882, May 2015.

[68] Hien Quoc Ngo, *Massive MIMO: Fundamentals and System Designs*, Linkoping University, Sweden, 2015.

[69] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, *Pilot Contamination and Precoding in Multi-Cell TDD Systems*, IEEE Transactions on Wireless Communications, vol. 10, no. 8, pp. 2640-2651, August 2011.

[70] Joao Vieira, Fredrik Rusek, Ove Edfors, Steffen Malkowsky, Liang Liu, and Fredrik Tufvesson, *Reciprocity Calibration for Massive MIMO: Proposal, Modeling, and Validation*, IEEE Transactions on Wireless Communications, vol. 16, no. 5, pp. 3042-3056, May 2017.

[71] Kaltenberger, F. et al., *Relative channel reciprocity calibration in MIMO/TDD systems*, Future Network and Mobile Summit, June 2010.

[72] Rogalin, R. et al., *Hardware-Impairment Compensation for Enabling Distributed Large-Scale MIMO*, Information Theory and Applications Workshop (ITA), February 2013.

[73] Rogalin, R. et al., *Scalable Synchronization and Reciprocity Calibration for Distributed Multiuser MIMO*, IEEE Transactions on Wireless Communications, vol. 13, no. 4, pp. 1815-1831, April 2014.

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

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

[76] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, *A coordinated approach to channel estimation in large-scale multiple-antenna systems*, IEEE Journal on Selected Areas in Communications, vol. 31, no. 2, pp. 264-273, February 2013.

[77] Mueller, R., Vehkaperae, M. and Cottatellucci, L., *Blind Pilot Decontamination*, IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, October 2014.

[78] Olakunle Elijah, Chee Yen Leow, Tharek Abdul Rahman, Solomon Nunoo, and Solomon Zakwoi Iliya, *A Comprehensive Survey of Pilot Contamination in Massive MIMO-5G System*, IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 905-923, November 2015.

[79] Felipe A. P. de Figueiredo et. al., *Channel estimation for massive MIMO TDD systems assuming pilot contamination and flat fading*, EURASIP Journal on Wireless Communications and Networking, vol. 2018, no. 14, pp. 1-10, January 2018.

[80] Caire, G. and Shamai, S., *On the achievable throughput of a multiantenna Gaussian broadcast channel*, IEEE Transactions on Information Theory, vol. 49, no. 7, pp. 1691-1706, July 2003.

[81] Windpassinger, C. et al., *A performance study of MIMO detectors*, IIEEE Transactions on Wireless Communications, vol. 5, no. 8, pp. 2004-2008, August 2006.

[82] Peel, C. B. et al., *A vector-perturbation technique for near-capacity multiantenna communication, part I: Channel inversion and regularization*, IEEE Transactions on Communications, vol. 53, no. 1, pp. 195-202, January 2005.

[83] Zu, K., de Lamare, R. C. and Haardt, M., *Generalized design of low-complexity block diagonalization type precoding algorithms for multiuser MIMO systems*, IEEE Transactions on Communications, vol. 61, no. 10, pp. 4232-4242, October 2013.

[84] Stephan Jaeckel, Leszek Raschkowski, Kai Borner, Lars Thiele, *QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials*, IEEE Transactions on Antennas and Propagation, vol. 62, issue 6, pp. 3242-3256, June 2014.

[85] S. Verdu, *Multiuser Detection*, Cambridge University Press, 1st Ed. 1998.

[86] R. C. de Lamare, *Massive MIMO Systems: Signal Processing Challenges and Research Trends*, URSI Radio Science Bulletin, 2013.

[87] Y. -C. Liang, E. Y. Cheu, L. Bai, and G. Pan, *On the relationship between MMSE-SIC and BI-GDFE receivers for large multiple-input multiple-output channels*, IEEE Transactions on Signal Processing, vol. 56, no. 8, pp. 3627-3637, August 2008.

[88] H. Zhao, H. Long, and W. Wang, *Tabu search detection for MIMO systems*, IEEE PIMRC, pp. 1-5, September 2007.

[89] Y. Sun, *A family of likelihood ascent search multiuser detectors: An upper bound of bit error rate and a lower bound of asymptotic multiuser efficiency*, IEEE Transactions on Communications, vol. 57, no. 6, pp. 1743-1752, June 2009.

[90] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge, UK, Cambridge University Press, 2003.

[91] D. Tse, and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge, UK, Cambridge University Press, 2005.

[92] J. Jaldén and B. Ottersten, *On the complexity of sphere decoding in digital communications*, IEEE Transactions on Signal Processing, vol. 53, no. 4, pp. 1474-1484, April 2005.

[93] L. Barbero and J. Thompson, *Fixing the complexity of the sphere decoder for MIMO detection*, IEEE Transactions on Wireless Communications, vol. 7, no. 6, pp. 2131-2142, June 2008.

[94] Monowar Hasan, Ekram Hossain, and Dusit Niyato, *Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches*, IEEE Communications Magazine, vol. 51, no. 6, pp. 86-93, June 2013.

[95] Md Shipon Ali, Ekram Hossain, and Dong In Kim, *LTE/LTE-A Random Access for Massive Machine-Type Communications in Smart Cities*, IEEE Communications Magazine, vol. 55, no. 1, pp. 76-83, January 2017.

[96] Han Seung Jang, Su Min Kim, Kab Seok Ko, Jiyoung Cha, and Dan Keun Sung, *Spatial Group Based Random Access for M2M Communications*, IEEE Communications Letters, vol. 18, no. 6, pp. 961-964, June 2014.

[97] Shree Krishna Sharma, and Xianbin Wang, *Towards Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions*, arXiv:1808.02924, August 2018.

[98] Felipe A. P. de Figueiredo et al., *Multi-stage Based Cross-Correlation Peak Detection for LTE Random Access Preambles*, Revista Telecomunicações, vol. 15, no. 2, pp. 1-7, October 2013.

[99] Shendi Wang, John S. Thompson, and Peter M. Grant, *Closed-Form Expressions for ICI/ISI in Filtered OFDM Systems for Asynchronous 5G Uplink*, IEEE Transactions on Communications, vol. 65, no. 11, pp. 4886-4898, November 2017.

[100] Abdoli J., Jia M., and Ma J., *Filtered OFDM: A New Waveform for Future Wireless Systems*, IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), July 2015.

[101] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, *Energy and spectral efficiency of very large multiuser MIMO systems*, IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436-1449, April 2013.

[102] A. Ashikhmin, T. L. Marzetta, L. Li, *Interference reduction in multi-cell massive MIMO systems I: Large scale fading precoding and decoding*, arXiv preprint arXiv:1411.4182, November 2014.

[103] Theodore S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd edition, Pearson Education Inc, 2002.

[104] Zheng Chen, and Emil Bjornson, *Channel Hardening and Favorable Propagation in Cell-Free Massive MIMO with Stochastic Geometry*, arXiv:1710.00395, June 2018.

[105] H. Q. Ngo, Erik G. Larsson, and Thomas L. Marzetta, *Aspects of Favorable Propagation in Massive MIMO*, European Signal Processing Conference (EUSIPCO), pp. 76-80, September 2014.

[106] H. Q. Ngo, and E. Larsson, *No downlink pilots are needed in TDD Massive MIMO*, IEEE Transactions on Wireless Communications, vol. 16, no. 5, pp. 2921-2935, May 2017.

[107] Jinho Choi, *Massive MIMO With Joint Power Control*, IEEE Wireless Communications Letters, vol. 3, no. 4, August 2014.

[108] Hong Yang, Thomas L. Marzetta, *Massive MIMO With Max-Min Power Control in Line-of-Sight Propagation Environment*, IEEE Transactions on Communications, vol. 65, no. 11, pp. 4685-4693, November 2017.

[109] M. Matthaiou, M. R. McKay, P. J. Smith, and J. A. Nossek, *On the condition number distribution of complex wishart matrices*, IEEE Transactions on Communications, vol. 58, no. 6, pp. 1705-1717, June 2010.

[110] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, *Massive MIMO performance evaluation based on measured propagation data*, IEEE Transactions on Wireless Communications, vol. 14, no. 7, pp. 3899-3911, July 2015.

[111] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, *Channel measurements for large antenna arrays*, International Symposium on Wireless Communication Systems (ISWCS), pp. 811-815, August 2012.

[112] A. O. Martinez, E. D. Carvalho, and J. O. Nielsen, *Towards very large aperture massive MIMO: A measurement based study*, Proceedings of IEEE Global Telecommunications Conference (GLOBECOM), pp. 281-286, December 2014.

[113] Alex Oliveras Martinez, Elisabeth De Carvalho, Jesper Odum Nielsen, *Massive MIMO Properties based on Measured Channels: Channel Hardening, User Decorrela-*

tion and Channel Sparsity, Asilomar Conference on Signals, Systems and Computers (ACSSC), November 2016.

[114] Pramod Viswanath and David Tse, *Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality*, IEEE Transactions on Information Theory, vol. 49, no. 8, August 2003.

[115] E. G. Larsson, and H. V. Poor, *Joint beamforming and broadcasting in massive MIMO*, IEEE Transactions on Wireless Communications , vol. 15, no. 4, pp. 3058-3070, April 2016.

[116] S. R. Saunders, and A. Aragon-Zavala, *Antennas and Propagation for Wireless Communication Systems*, John Wiley & Sons, Chichester, UK, 2nd edition, 2007.

[117] Kamil Senel, and Erik G. Larsson, *Grant-Free Massive MTC-Enabled Massive MIMO: A Compressive Sensing Approach*, arXiv preprint arXiv:1806.10061, June 2018.

[118] Tarik Taleb, Andreas Kunz, *Machine type communications in 3GPP networks: potential, challenges, and solutions*, IEEE Communications Magazine, vol. 50, no. 3, March 2012.

[119] Liang Liu, and Wei Yu, *Massive Connectivity With Massive MIMO - Part I: Device Activity Detection and Channel Estimation*, IEEE Transactions on Signal Processing, vol. 66, no 11, pp. 2933-2946, June 2018.

[120] Kamil Senel, Emil Bjornson, Erik G Larsson, *Human and Machine Type Communications can Coexist in Uplink Massive MIMO Systems*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6613-6617, April 2018.

[121] Rasha Al Khansa, Jean J. Saade, Hassan A. Artail, and Mohamad Assaad, *A small cell approach to optimizing the coverage of MTC systems with massive MIMO and random access using stochastic geometry*, IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), October 2017.

[122] 3GPP, *TS 22.368 V10.1.0 - Service Requirements for Machine-Type Communications*, June 2010.

[123] R. L. Frank, S. A. Zadoff and R. Heimiller, *Phase shift pulse codes with good periodic correlation properties*, IRE Transactions on Information Theory, vol. 8, no. 6, pp. 381-382, October 1962.

[124] David C. Chu, *Polyphase codes with good periodic correlation properties*, IEEE Transactions on Information Theory, vol. 18, no. 4, pp. 531-532, July 1972.

[125] Wallis, J. S., *On the existence of Hadamard matrices*, Journal of Combinatorial Theory. vol. 21, no. 2, pp. 188–195, September 1976.

[126] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, New York, NY, USA: Pearson Education, 1993.

[127] Pallaviram Sure, and Chandra Mohan Bhuma, *A survey on OFDM channel estimation techniques based on denoising strategies*, Engineering Science and Technology, an International Journal, vol. 20, no. 2, pp. 629-636, April 2017.

[128] Edmund Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, AMS Chelsea Publishing, 1909.

[129] J. Hoydis, S. ten Brink, and M. Debbah, *Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?*, IEEE Journal on Selected Areas in Communications, vol. 31, no. 2, pp. 160-171, February 2013.

[130] 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 Transactions on Communications, vol. 61, no. 6, pp. 2350-2361, June 2013.

[131] N. Shariati, E. Bjornson, M. Bengtsson, and M. Debb, *Low-Complexity Polynomial Channel Estimation in Large-Scale MIMO with Arbitrary Statistics*, IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 815-830, October 2014.

[132] K. Mawatwal, D. Sen and R. Roy, *A Semi-Blind Channel Estimation Algorithm for Massive MIMO Systems*, IEEE Wireless Communications Letters, vol. 6, no. 1, pp. 70-73, February 2017.

[133] Ke Li, Xiaoqin Song, M. Omair Ahmad, and M. N. S. Swamy, *An Improved Multicell MMSE Channel Estimation in a Massive MIMO System*, International Journal of Antennas and Propagation, vol. 2014, May 2014.

[134] Ang Yang, Zunwen He, Chengwen Xing, Zesong Fei, and Jingming Kuang, *The Role of Large-Scale Fading in Uplink Massive MIMO Systems*, IEEE Transactions on Vehicular Technology, vol. 65, no. 1, pp. 477-483, January 2016.

[135] Ko-Feng Chen, Yen-Cheng Liu, Yu T Su, *On composite channel estimation in wireless massive MIMO systems*, IEEE Globecom Workshops, pp. 135-139, September 2013.

[136] E. Bjornson, E. Jorswieck, and B. Ottersten, *Impact of Spatial Correlation and Precoding Design in OSTBC MIMO Systems*, IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3578-3589, November 2010.

[137] A. Chockalingam, and B. S. Rajan, *Large MIMO Systems*, Cambridge, U.K. Cambridge Univ. Press, 2013.

[138] T. L. Marzetta, *How much training is required for multiuser MIMO?*, in Proc. 40th Asilomar Conf. Signals, Syst., Comput. (ACSSC), Pacific Grove, CA, USA, pp. 359-363, October 2006.

[139] 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, June 2010.

[140] Emil Bjornson, Erik G. Larsson, and Mérouane Debbah, *Massive MIMO for Maximal Spectral Efficiency: How Many Users and Pilots Should Be Allocated?*, IEEE Transactions on Wireless Communications, vol. 15, no. 2, pp. 1293-1308, February 2016.

[141] C. Radhakrishna Rao, *Linear Statistical Inference and Its Applications*, 2nd. ed., Wiley, New York, 1973.

[142] Fisher, R.A. *On the mathematical foundations of theoretical statistics*, Philosophical Transactions of the Royal Society, vol. 222, no. 594-604, pp. 309-368, 1922.

[143] Amin Khansefid, Hlaing Minn, *On Channel Estimation for Massive MIMO With Pilot Contamination*, IEEE Communications Letters, vol. 19, no. 9, pp. 1660-1663, September 2015.

[144] IEEE 754-2008, *IEEE 754–2008: Standard for Floating-Point Arithmetic*, Institute of Electrical and Electronics Engineers, August 2008.

[145] Patrick Billingsley, *Probability and Measure*, Third Edition, New York: Wiley, pp. 231-240, ISBN 0-471-00710-2, 1995.

[146] Wolfram Research, Inc., *Mathematica*, Version 11.3, Champaign, IL (2018).

[147] Emil Bjornson et. al., *Massive MIMO is a Reality - What is Next? Five Promising Research Directions for Antenna Arrays*, submitted to Digital Signal Processing, arXiv:1902.07678, February 2019.

[148] Ricardo Tadashi Kobayashi, and Taufik Abrão, *Theoretical error for asynchronous multi-user large-scale MIMO channel estimation*, IET Communications, vol. 11, no. 1, pp. 17-24, December 2016.

[149] Jun Fang, Xingjian Li, Hongbin Li, and Feifei Gao, *Low-Rank Covariance-Assisted Downlink Training and Channel Estimation for FDD Massive MIMO Systems*, IEEE Trans. on Wireless Commun., vol. 16, no. 3, March 2017.

[150] Mojtaba Soltanalian, Mohammad Mahdi Naghsh, Nafiseh Shariati, Petre Stoica, and Babak Hassibi, *Training Signal Design for Correlated Massive MIMO Channel Estimation*, IEEE Trans. on Wireless Commun., vol. 16, no. 2, February 2017.

[151] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, *Pilot beam pattern design for channel estimation in massive MIMO systems*, IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 787-801, October 2014.

[152] K. T. Truong and R. W. Heath, *Effects of Channel Aging in Massive MIMO Systems*, Journal of Communications and Networks, vol. 15, no. 4, pp. 338-351, August 2013.

[153] Heshani Gamage, Nandana Rajatheva, and Matti Latva-aho, *Performance and PAPR Analysis of Single-Carrier Massive MIMO Systems with Channel Imperfections*, IEEE Vehicular Technology Conference (VTC-Fall), September 2017.

[154] Bernard Sklar, *Rayleigh Fading Channels in Mobile Digital Communication Systems Part I: Characterization*, IEEE Communications Magazine, vol. 35, no. 7, pp. 90-100, September 1997.

[155] W. H. Tranter, K. S. Shanmugan, T. S. Rappaport and K. L. Kosbar, *Principles of Communication Systems Simulation with Wireless Applications*, Prentice Hall, ISBN 0-13-494790-8, 2004.

[156] N. Jindal, *MIMO broadcast channels with finite-rate feedback*, IEEE Transactions on Information Theory, vol. 52, no. 11, pp. 5045-5060, November 2006.

[157] A. Jeffrey and D. Zwillinger, *Table of Integrals, Series, and Products*, New York, NY, USA: Academic, 2007.

[158] Emil Bjornson, Jakob Hoydis, Luca Sanguinetti, *Massive MIMO has Unlimited Capacity*, IEEE Trans. Wireless Commun., vol. 17, no. 1, pp. 574-590, January 2018.

[159] George Marsaglia, *Ratios of Normal Variables*, Journal of Statistical Software, vol. 16, no. 4, May 2006.

Appendix A – List of Publications

[1] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Renato R. Lopes, and Joao Paulo Miranda, *On the Application of Massive MU-MIMO in the Uplink of Machine Type Communication Systems*, International Workshop on Telecommunications (IWT), June 2015.

[2] João P. Miranda, Arman Farhang, Nicola Marchetti, Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, and Fabrício Figueiredo, *On massive MIMO and its applications to machine type communications and FBMC-based networks*, EAI Endorsed Transactions on Ubiquitous Environments, vol. 2, no. 5, July 2015.

[3] Felipe A. P. de Figueiredo, Fabiano Mathilde, Fabricio Santos, Fabbryccio A. C. M. Cardoso, and Gustavo Fraidenraich, *On channel estimation for massive MIMO with pilot contamination and multipath fading channels*, IEEE Latin-American Conference on Communications (LATINCOM), October 2016.

[4] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman, and Gustavo Fraidenraich, *Channel Estimation for Massive MIMO TDD Systems Assuming Pilot Contamination and Frequency-Selective Fading*, IEEE Access, vol. 5, no. 9, pp. 17733-17741, September 2017.

[5] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman, and Gustavo Fraidenraich, *Channel estimation for massive MIMO TDD systems assuming pilot contamination and flat fading*, EURASIP Journal on Wireless Communications and Networking, vol. 2018, no. 14, January, 2018.

[6] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman, Gustavo Fraidenraich, *On the Application of Massive MIMO Systems to Machine Type Communications*, IEEE Access, vol. 7, no. 12, pp. 2589-2611, December 2018.

[7] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, and Gustavo Fraidenraich, *On the distribution of an effective channel estimator for multi-cell massive MIMO*, IEEE Access, vol. 7, no. 1, pp. 114508-114519, December 2019.

[8] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman, and Gustavo Fraidenraich, *Large-Scale Antenna Systems and Massive Machine Type Communications*, (under revision, submitted to Annals of Telecommunications).

[9] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, and Gustavo Fraidenraich, *Performance Analysis of Large-Scale MU-MIMO With a Simple and Effective Chanel Estimator*, submitted to IEEE Latin-American Conference on Communications (LATINCOM 2019).

Appendix B – Proof of $\lim_{M\to\infty} \hat{\beta}_k = \beta_k$ defined in Eq. (3.39).

For the proof of (3.39), we need the following Lemmas.

Lemma 1. *Let $z \sim \mathcal{CN}(0,a)$, where $z = u + jv$, then $\mathbb{E}\left[|z|^4\right] = 2a^2$.*

Proof. In order to prove this, we first rewrite $\mathbb{E}\left[|z|^4\right]$ as

$$\mathbb{E}\left[|z|^4\right] = \frac{1}{\pi a}\int_{\mathbb{C}} |z|^4 e^{-\frac{|z|^2}{a}} dz. \tag{B.1}$$

Next, taking the Jacobian determinant of the transformation $r = \sqrt{u^2+v^2}$ and $\theta = \arctan\left(\frac{v}{u}\right)$ results in

$$\begin{aligned} |J| = \left|\frac{\partial(r,\theta)}{\partial(u,v)}\right| = \begin{vmatrix} \frac{\partial r}{\partial u} & \frac{\partial r}{\partial v} \\ \frac{\partial \theta}{\partial u} & \frac{\partial \theta}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{u}{\sqrt{u^2+v^2}} & \frac{v}{\sqrt{u^2+v^2}} \\ -\frac{v}{u^2+v^2} & \frac{u}{u^2+v^2} \end{vmatrix} \\ = \frac{1}{\sqrt{u^2+v^2}} = \frac{1}{r}. \end{aligned} \tag{B.2}$$

Therefore, the forth moment of z expressed in terms of r and θ is defined by

$$\begin{aligned} \mathbb{E}\left[r^4\right] = \mathbb{E}\left[|z|^4\right]\left|\frac{\partial(r,\theta)}{\partial(u,v)}\right| \\ = \frac{1}{\pi}\int_{-\pi}^{\pi} d\theta \frac{1}{a}\int_0^\infty r^5 e^{-\frac{r^2}{a}} dr \\ = \frac{2}{a}\int_0^\infty r^5 e^{-\frac{r^2}{a}} dr \\ = 2a^2. \end{aligned} \tag{B.3}$$

■

Lemma 2. *Let $\boldsymbol{z} = [z_1, \ldots, z_M]^T \in \mathbb{C}^{M\times 1}$ be a complex random vector with distribution $\boldsymbol{z} \sim \mathcal{CN}(\boldsymbol{0}_M, a\boldsymbol{I}_M)$. Then $\mathbb{E}\left[\|\boldsymbol{z}\|^4\right] = a^2 M(M+1)$.*

Proof. We start the proof by expanding $\mathbb{E}\left[\|\mathbf{z}\|^4\right]$ as

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{z}\|^4\right] = \mathbb{E}\left[(\mathbf{z}^H\mathbf{z})^2\right] = \mathbb{E}\left[\left(\sum_{m=1}^{M} |z_m|^2\right)^2\right] \\ = \sum_{m=1}^{M}\sum_{m'=1}^{M} \mathbb{E}\left[|z_m|^2|z_{m'}|^2\right] \\ = \sum_{m=1}^{M} \mathbb{E}\left[|z_m|^4\right] + \sum_{m=1}^{M}\sum_{m'=1,\, m'\neq m}^{M} \mathbb{E}\left[|z_m|^2\right]\mathbb{E}\left[|z_{m'}|^2\right] \\ = a^2 M(M+1), \end{aligned} \tag{B.4}$$

where in the last equality we used Lemma 1. Lemma 3, which is defined below, can be used to find moments of any order.

■

Lemma 3. *Let* $\boldsymbol{z} = [z_1, \ldots, z_M]^T \in \mathbb{C}^{M\times 1}$ *be a complex random vector with distribution* $\boldsymbol{z} \sim \mathcal{CN}(\boldsymbol{0}_M, a\boldsymbol{I}_M)$. *Then* $\mathbb{E}\left[\|\boldsymbol{z}\|^{2k}\right] = a^k \prod_{i=0}^{k-1}(M+i)$, *where* $\{k \in \mathbb{Z} \mid k > 0\}$.

Proof. Given the pdf of z

$$f(\mathbf{z}) = \frac{1}{(\pi a)^M} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right). \tag{B.5}$$

By using the following identity

$$\frac{1}{(\pi a)^M} \int_{\mathbb{C}} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z} = 1, \tag{B.6}$$

which can be rewritten as

$$\int_{\mathbb{C}} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z} = (\pi a)^M, \tag{B.7}$$

and next deriving both sides of the identity w.r.t. a, we find

$$\int_{\mathbb{C}} \frac{d^k}{d^k a}\left[\exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right)\right] d\mathbf{z} = \frac{d^k}{d^k a}\left[(\pi a)^M\right]. \tag{B.8}$$

After applying the derivation to (B.8), it can be rewritten as

$$\int_{\mathbb{C}} \|\mathbf{z}\|^{2k} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z} = \pi^M (M)^{(k)} a^{M+k}, \tag{B.9}$$

where $(M)^{(k)} = \prod_{i=0}^{k-1}(M+i)$. Next, by reorganizing (B.9), we find

$$\underbrace{\frac{1}{(\pi a)^M} \int_{\mathbb{C}} \|\mathbf{z}\|^{2k} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z}}_{\mathbb{E}\left[\|\mathbf{z}\|^{2k}\right]} = a^k (M)^{(k)}, \tag{B.10}$$

which concludes the proof. ■

Lemma 4. *Let* $\boldsymbol{z} = [z_1, \ldots, z_M]^T \in \mathbb{C}^{M\times 1}$ *be a complex random vector with distribution* $\mathcal{CN}(\boldsymbol{0}_M, a\boldsymbol{I}_M)$. *Then*

$$\lim_{M\to\infty} \frac{\|\boldsymbol{z}\|^2}{M} \overset{\text{a.s.}}{=} a. \tag{B.11}$$

Proof. As M increases, $\frac{\mathbf{z}^H\mathbf{z}}{M}$ becomes more and more deterministic, and consequently, its variance must be zero when $M \to \infty$. Therefore, one way to prove the convergence in

(B.11) is to show that

$$\begin{aligned} &\lim_{M\to\infty} \operatorname{var}\left[\frac{\|\mathbf{z}\|^2}{M}\right] \\ &= \lim_{M\to\infty} \frac{1}{M^2}\operatorname{var}\left[\|\mathbf{z}\|^2\right] \\ &= \lim_{M\to\infty} \frac{1}{M^2}\left\{\mathbb{E}\left[\|\mathbf{z}\|^4\right] - \left(\mathbb{E}\left[\|\mathbf{z}\|^2\right]\right)^2\right\}. \end{aligned} \tag{B.12}$$

Applying Lemma 2 to (B.12) and knowing that $\mathbb{E}\left[\|\mathbf{z}\|^2\right] = aM$, we find that

$$\lim_{M\to\infty} \operatorname{var}\left[\frac{\|\mathbf{z}\|^2}{M}\right] = a^2 \lim_{M\to\infty} \frac{1}{M} \overset{\text{a.s.}}{=} 0. \tag{B.13}$$

This Lemma is in accordance with the Law of large numbers [104, 106].

For the proof of (3.39) we take into account the distribution of $\mathbf{y}_k$ defined in (3.3) and using Lemma 4, after simple computations, we complete the proof.

■

Appendix C – Proof of the exact MSE per antenna defined in Eq. (3.43).

For the proof of the mean-square estimation error of the proposed channel estimator, we need to define the following Lemmas.

Lemma 5. *If* $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_x^2 \boldsymbol{I}_M)$ *and* $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_y^2 \boldsymbol{I}_M)$ *are independent and* $\frac{\boldsymbol{x}^H}{\|\boldsymbol{x}\|}\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \triangleq Re^{j\theta}$*, therefore,* θ *is uniformly distributed in the range* $[-\pi, \pi]$ *and the pdf of* R *is defined as*

$$f_R(r) = 2Mr(1-r^2)^{M-1}, \;\; 0 \leq r \leq 1. \tag{C.1}$$

Proof. The circular symmetry of $\mathbf{x}$ and $\mathbf{y}$ results in the uniform distribution of θ. The random variable $Z = \left|\frac{\mathbf{x}^H}{\|\mathbf{x}\|}\frac{\mathbf{y}}{\|\mathbf{y}\|}\right|^2$ exhibits a Beta pdf $f_Z(z) = M(1-z)^{M-1}, \;\; z \in [0,1]$ [156]. Finally, the transformation of random variable $R = \sqrt{Z}$ yields (C.1). ∎

Lemma 6. *If a random variable* R *has pdf defined as in* (C.1)*,* θ *is uniformly distributed within the range* $[-\pi, \pi]$ *and they are independent, therefore the pdf of the random variable* $W \triangleq R\cos(\theta)$ *is given by*

$$f_W(w) = \frac{M}{\pi} B\left(\frac{1}{2}, M\right)\left(1-w^2\right)^{M-\frac{1}{2}}, \;\; |w| \leq 1. \tag{C.2}$$

Proof. Starting from the uniform distribution of θ, we have that

$$\mathbb{P}\left(\cos(\theta) \leq \frac{w}{r}\right) = \begin{cases} 0, & \frac{w}{r} < -1 \\ 1 - \frac{1}{\pi}\cos^{-1}\left(\frac{w}{r}\right), & -1 \leq \frac{w}{r} \leq 1 \\ 1, & \frac{w}{r} > 1 \end{cases} \tag{C.3}$$

Next, the cumulative distribution function (cdf) of the random variable W is defined as $F_W(w) = \mathbb{P}\left(\cos(\theta) \leq w\right) = \int_0^1 \mathbb{P}\left(\cos(\theta) \leq \frac{w}{r}\right) f_R(r)dr$

$$F_W(w) = \begin{cases} 0, & w < -1 \\ \xi(w), & -1 \leq w \leq 0 \\ \xi(w) + \int_0^w f_R(r)dr, & 0 \leq w \leq 1 \\ 1, & w > 1 \end{cases} \tag{C.4}$$

where $\xi(w) \triangleq \int_{|w|}^1 \left(1 - \frac{1}{\pi}\cos^{-1}\left(\frac{w}{r}\right)\right) f_R(r)dr$. Therefore, we find that $f_W(w) = \frac{dF_W(w)}{dw} = \int_{|w|}^1 \frac{2Mr(1-r^2)^{M-1}}{\pi\sqrt{r^2-w^2}}dr, \;\; |w| < 1$. Next, by changing the variable z as $z = r^2 - w^2$ and using [157], we complete the proof and find (C.2). ∎

Lemma 7. *If the random variables* $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}_M, \boldsymbol{I}_M)$ *and* $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}_M, \boldsymbol{I}_M)$ *are independent, therefore,* $U = \frac{\|\boldsymbol{x}\|}{\|\boldsymbol{y}\|}$ *has its pdf defined by*

$$f_U(u) = \frac{2\Gamma(2M)}{(\Gamma(M))^2} \frac{u^{2M-1}}{(u^2+1)^{2M}}, \quad u > 0. \tag{C.5}$$

Proof. We start by recalling that $\|\mathbf{x}\|^2$ and $\|\mathbf{y}\|^2$ are central Chi-square random variables with the following pdf: $f_V(v) = \frac{v^{M-1}}{\Gamma(M)} e^{-v}$. Next, by using the independence of $\|\mathbf{x}\|^2$ and $\|\mathbf{y}\|^2$, the cdf of U^2 is found and then by differentiating it we find the pdf of U^2 as being defined by $f_Z(z) = \frac{\Gamma(2M)}{(\Gamma(M))^2} \frac{z^{M-1}}{(z+1)^{2M}}, z > 0$. Finally, by applying the square root variable transformation to U^2, we find (C.5). ■

Lemma 8. *If the random variables* $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_x^2 \boldsymbol{I}_M)$ *and* $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_y^2 \boldsymbol{I}_M)$ *are independent, therefore*

$$\mathbb{E}\left\{\frac{(\boldsymbol{x}+\boldsymbol{x})^H \boldsymbol{x}}{\|\boldsymbol{x}+\boldsymbol{y}\|^2}\right\} = \int_0^\infty \int_{-1}^1 \frac{(ku+w) f_U(u) f_W(w)}{ku + \frac{1}{ku} + 2w} dw du, \tag{C.6}$$

where $k \triangleq \sqrt{\frac{\sigma_x^2}{\sigma_y^2}}$*, and* $f_W(w)$ *and* $f_U(u)$ *are defined in* (C.2) *and* (C.5)*, respectively.*

Proof. We start this proof by expanding and dividing both the numerator and denominator of the left-hand part of (C.6) by $\|\mathbf{x}\|\|\mathbf{y}\|$. This way, that term can be re-written as

$$\mathbb{E}\left\{\frac{(\mathbf{x}+\mathbf{y})^H \mathbf{x}}{\|\mathbf{x}+\mathbf{y}\|^2}\right\} = \mathbb{E}\left\{\frac{kU + Re^{-j\theta}}{kU + \frac{1}{kU} + 2W}\right\}, \tag{C.7}$$

where $\frac{\|\mathbf{x}\|}{\|\mathbf{y}\|} \triangleq kU$, $\frac{\mathbf{x}^H}{\|\mathbf{x}\|} \frac{\mathbf{y}}{\|\mathbf{y}\|} \triangleq Re^{j\theta}$, and $W \triangleq R\cos(\theta)$. Note that U, R and θ are independent random variables. Initially, we find that the expected value of the imaginary part of the left-hand side of (C.7) is equal to zero when we condition it on U and R, and average it over θ. Therefore, (C.7) becomes $\mathbb{E}\left\{\frac{kU+W}{kU+\frac{1}{kU}+2W}\right\}$. Next, by applying Lemmas 6 and 7 to (C.7) and doing a direct calculation of the expectation in (C.7) over the pdfs, $f_U(u)$ and $f_W(w)$, results in (C.6), which completes the proof. ■

Proof of the mean-square estimation error, η_k^{prop}

For the proof of the mean-square estimation error, we first expand it as

$$\begin{aligned}\eta_k^{\text{prop}} = \frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_k^{\text{prop}}\|^2\right] + \frac{1}{M}\mathbb{E}\left[\|\mathbf{g}_k\|^2\right] \\ -\frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_k^{\text{prop}})^H \mathbf{g}_k\right]\right],\end{aligned} \tag{C.8}$$

and find these three expectations. The first expectation can be expanded as

$$\begin{aligned}\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_k^{\text{prop}}\|^2\right] &= \frac{1}{M}\mathbb{E}\left[\left(1 - \frac{M}{\rho}\frac{1}{\|\mathbf{y}_k\|^2}\right)^2 \|\mathbf{y}_k\|^2\right] \\ &= \frac{1}{M}\mathbb{E}\left[\|\mathbf{y}_k\|^2\right] - \frac{2}{\rho} + \frac{M}{\rho^2}\mathbb{E}\left[\frac{1}{\|\mathbf{y}_k\|^2}\right],\end{aligned} \tag{C.9}$$

where the first term $\frac{1}{M}\mathbb{E}\left[\|\mathbf{y}_k\|^2\right] = \beta_k + \frac{1}{\rho}$ and in order to find the last term we use the fact that $\frac{1}{\|\mathbf{y}_k\|^2}$ has an Inverse Gamma distribution, $\Gamma^{-1}(M, \beta_k + \frac{1}{\rho})$, with mean, $\mathbb{E}\left[\frac{1}{\|\mathbf{y}_k\|^2}\right] = \frac{1}{(\beta_k+\frac{1}{\rho})(M-1)}$, therefore, we have the first expectation defined as

$$\begin{aligned} \frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_k^{\text{prop}}\|^2\right] &= \frac{(\rho\beta_k+1)(\rho\beta_k-1)(M-1)+M}{\rho(\rho\beta_k+1)(M-1)} \\ &= \frac{\rho\beta_k^2}{(\rho\beta_k+1)} + \frac{1}{\rho(\rho\beta_k+1)(M-1)}. \end{aligned} \tag{C.10}$$

Next, for the second expectation, we recall that $\|\mathbf{g}_k\|^2$ has a Gamma distribution, $\Gamma(M, \beta_k)$ and therefore, $\frac{1}{M}\mathbb{E}\left[\|\mathbf{g}_k\|^2\right] = \beta_k$.

Finally, in order to find the third expectation, we first expand it as

$$\begin{aligned} &-\frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_k^{\text{prop}})^H\mathbf{g}_k\right]\right] \\ &= -\frac{2}{M}\mathbb{E}\left[\Re\left\{\left(1-\frac{M}{\rho}\frac{1}{\|\mathbf{y}_k\|^2}\right)\mathbf{y}_k^H\mathbf{g}_k\right\}\right] \\ &= -\frac{2}{M}\mathbb{E}\left[\Re\left\{\mathbf{y}_k^H\mathbf{g}_k\right\}\right] + \frac{2}{\rho}\mathbb{E}\left[\Re\left\{\frac{\mathbf{y}_k^H\mathbf{g}_k}{\|\mathbf{y}_k\|^2}\right\}\right]. \end{aligned} \tag{C.11}$$

After simple calculations we find that the first term in (C.11) is given by $-\frac{2}{M}\mathbb{E}\left[\Re\left\{\mathbf{y}_k^H\mathbf{g}_k\right\}\right] = -2\beta_k$. In order to find the second term in (C.11), we apply Lemma 8 to it with $\mathbf{x} = \mathbf{g}_k$ and $\mathbf{y} = \mathbf{y}_k$, where the distribution of $\mathbf{y}_k$ is defined in (3.3). Next, for the purpose of making the boundary in the integral finite, we apply the following change of variable, $t = \frac{1}{1+u^2}$, and then, we define this integral as θ_k. After substituting each one of the three expectations back in the expansion of the mean-square estimation error, η_k^{prop}, in (C.8), we conclude the proof.

Appendix D – Proof of the MSE between $\hat{\mathbf{g}}_k^{\text{prop}}$ and $\hat{\mathbf{g}}_k^{\text{MMSE}}$ defined in Eq. (3.46).

In this appendix we present proof for Remark 9. We start with the following expansion

$$\frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_k^{\text{prop}} - \hat{\mathbf{g}}_k^{\text{MMSE}}\|^2] = \frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_k^{\text{prop}}\|^2\right] + \frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_k^{\text{MMSE}}\|^2\right] - \frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_k^{\text{prop}})^H \hat{\mathbf{g}}_k^{\text{MMSE}}\right]\right]. \tag{D.1}$$

Next, we calculate the three different expectations in (D.1). First, from Appendix C we know that $\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_k^{\text{prop}}\|^2\right] = \gamma_k + \epsilon_k$. Next, by recalling that the distribution of $\hat{\mathbf{g}}_k^{\text{MMSE}}$ is defined in (3.15), and using the fact that $\|\hat{\mathbf{g}}_k^{\text{MMSE}}\|^2$ has a Gamma distribution, $\Gamma\left(M, \frac{\rho\beta_k^2}{1+\rho\beta_k}\right)$, then we have that $\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_k^{\text{MMSE}}\|^2\right] = \frac{\rho\beta_k^2}{1+\rho\beta_k} = \gamma_k$. For the last expectation, after substituting $\hat{\mathbf{g}}_k^{\text{prop}}$ and $\hat{\mathbf{g}}_k^{\text{MMSE}}$ in the last term of (D.1) we find

$$\begin{aligned} &-\frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_k^{\text{prop}})^H \hat{\mathbf{g}}_k^{\text{MMSE}}\right]\right] \\ &= -\frac{2}{M}\frac{\beta_k}{\beta_k + \frac{1}{\rho}}\left\{\mathbb{E}\left[\|\mathbf{y}_k\|^2\right] - \frac{M}{\rho}\right\} \\ &= -2\frac{\rho\beta_k^2}{1+\rho\beta_k} = -2\gamma_k, \end{aligned} \tag{D.2}$$

where we have used $\mathbb{E}\left[\|\mathbf{y}_k\|^2\right] = M(\beta_k + \frac{1}{\rho})$ in the last equality. Finally, after substituting the three found expectations back in (D.1), we conclude the proof.

Appendix E – Proof of the approximated MSE per antenna defined in Eq. (3.47).

For the proof of the approximated mean-square estimation error per antenna of the proposed estimator, we need to define the following Lemma.

Lemma 9. *Let μ_X and μ_Y be the expectations of X and Y, σ_Y^2 be the variance of Y, and σ_{XY} be their covariance. Then the expectation, $\mathbb{E}\{X/Y\}$, can be approximated by*

$$\mathbb{E}\left\{\frac{X}{Y}\right\} \approx \frac{\mu_X}{\mu_Y} - \frac{\sigma_{XY}}{\mu_Y^2} + \frac{\mu_X}{\mu_Y^3}\sigma_Y^2. \tag{E.1}$$

Proof. For a function that depends on two variables, x and y, the second order Taylor expansion series about the point (a, b) is given by

$$\begin{aligned} g(x,y) = g(a,b) + g_x(a,b)(x-a) + g_y(a,b)(y-b) + \\ + \frac{1}{2!}(g_{xx}(a,b)(x-a)^2 + 2g_{xy}(a,b)(x-a)(y-b) + \\ + g_{yy}(a,b)(y-b)^2), \end{aligned} \tag{E.2}$$

where the subscripts denote the respective partial derivatives. The partial derivatives are defined by $g_y = -X/Y^2$, $g_{yy} = 2X/Y^3$, $g_x = 1/Y$, $g_{xx} = 0$, and $g_{xy} = -1/Y^2$. Applying the derivatives into (E.2), the second order Taylor expansion of $g(X, Y) = X/Y$ around the mean point (μ_X, μ_Y), the following is obtained

$$\begin{aligned} \frac{X}{Y} \approx \frac{\mu_x}{\mu_y} - \frac{\mu_x}{\mu_y^2}(Y-\mu_y) + \frac{1}{\mu_y}(X-\mu_x) + \\ + \frac{1}{2!}\left(\frac{2\mu_x}{\mu_y^3}(Y-\mu_y)^2 - \frac{2}{\mu_y^2}(Y-\mu_y)(X-\mu_x)\right). \end{aligned} \tag{E.3}$$

Finally, applying the expectation operator, $\mathbb{E}\{.\}$, to (E.3) concludes the proof.

■

With the purpose of finding a more tractable expression for the mean-square error per antenna of the proposed channel estimator we derive an approximation to the expectation of the ratio of random variables in the last part of (C.11), *i.e.*, the term defined as θ_k.

It is possible to approximate the moments of a function $g(X, Y)$ using Taylor series expansions, provided that g is sufficiently differentiable and that the moments of

X and Y are finite. Therefore, applying Lemma 9 to

$$\theta_k = \mathbb{E}\left[\Re\left\{\frac{\mathbf{y}_k^H \mathbf{g}_k}{\|\mathbf{y}_k\|^2}\right\}\right] = \mathbb{E}\left[\Re\left\{\frac{(\mathbf{g}_k + \mathbf{w'})^H \mathbf{g}_k}{\|\mathbf{g}_k + \mathbf{w'}\|^2}\right\}\right], \tag{E.4}$$

we are able to find an approximation to θ_k, which is defined as

$$\theta_k \approx \frac{\rho\beta_k}{1 + \rho\beta_k}. \tag{E.5}$$

Note that the approximation does not depend on M but only on ρ. The proof is concluded by substituting (E.5) into the expansion of the mean-square error given in (D.1).

Appendix F – Proof of the approximated MSE per antenna defined in Eq. (4.18).

For the proof of the approximate MSE of the proposed estimator in Chapter 4 we need to present a few Lemmas.

Lemma 10. *If* $W \sim \mathcal{CN}(0, \sigma^2)$ *then* $\|W\|^2 \sim \Gamma(1, \sigma^2)$.

Proof. We know that $W = X + jY$ and consequently $\|W\|^2 = X^2 + Y^2$, where X, Y are i.i.d. random variables with distribution $\mathcal{N}(0, \sigma^2/2)$. Now, if we make $Z = X^2 + Y^2$, then using the joint pdf of X and Y we have that $\mathbb{P}(Z \leq z)$ is defined by

$$\mathbb{P}(X^2 + Y^2 \leq z) = \frac{1}{\pi\sigma^2} \int_{x^2+y^2 \leq z} e^{-\frac{(x^2+y^2)}{\sigma^2}} \, dxdy. \tag{F.1}$$

Next, switching to polar coordinates we get:

$$\mathbb{P}(Z \leq z) = \frac{1}{\pi\sigma^2} \int_0^{2\pi} \int_0^{\sqrt{z}} r e^{-\frac{r^2}{\sigma^2}} dr d\theta = \frac{2}{\sigma^2} \int_0^{\sqrt{z}} r e^{-\frac{r^2}{\sigma^2}} dr. \tag{F.2}$$

Now if we set $u = r^2$ then we get

$$\mathbb{P}(Z \leq z) = \frac{1}{\sigma^2} \int_0^{z} e^{-\frac{u}{\sigma^2}} du, \tag{F.3}$$

so Z is exponentially distributed with rate parameter $\lambda = 1$.

Finally, comparing the pdf given above with the Gamma distribution pdf one can notice that if the shape parameter k is set to 1 and scale parameter θ is set to σ^2 it becomes the exponential pdf, concluding the proof. ∎

Lemma 11. *If* $X_m \sim \mathcal{CN}(0, \sigma^2)$ $\forall m$ *are independent, then* $\sum_{m=1}^{M} |X_m|^2 \sim \Gamma(M, \sigma^2)$.

Proof. From Lemma 10 we know that each $Z_m = \|X_m\|^2 \sim \Gamma(1, \sigma^2)$. We also know that Z_m is independent for all m. Therefore, each Z_m has the following characteristic function:

$$\varphi_Z(t) = (1 - j\sigma^2 t)^{-1}. \tag{F.4}$$

Next, knowing that the characteristic function of the sum of independent random variables is the product of their individual characteristic functions leads to

$$\varphi_{Z_1+Z_2+\cdots+Z_M}(t) = \varphi_{Z_1}(t)\varphi_{Z_2}(t)\cdots\varphi_{Z_M}(t) = (1 - j\sigma^2 t)^{-M}. \tag{F.5}$$

Eq. (F.5) defines the characteristic function of a Gamma-distributed random variable with scale parameter $\theta = \sigma^2$ and shape parameter $k = M$, and therefore concluding the proof. ■

Lemma 12. *If* $X \sim \Gamma(k, \theta)$ *and* $\frac{1}{X} \sim \Gamma^{-1}(k, \theta)$*, i.e., the Inverse-Gamma distribution, then* $\mathbb{E}\left\{\frac{1}{X}\right\} = \frac{1}{\theta(k-1)}$.

Proof. We start this proof by defining the expectation of $Y = \frac{1}{X}$, which is given by

$$\mathbb{E}\{Y\} = \int_0^\infty y \frac{1}{\Gamma(k)\theta^k}\left(\frac{1}{y}\right)^{k+1} e^{-\frac{1}{\theta}\left(\frac{1}{y}\right)} dy. \tag{F.6}$$

Next we apply the following change of variable, $z = \frac{1}{y}$, to (F.6), which becomes

$$\mathbb{E}\{Z\} = \frac{1}{\Gamma(k)\theta^k}\int_0^\infty z^{k-2} e^{-\frac{1}{\theta}z} dz. \tag{F.7}$$

By using the following identity

$$\int_0^\infty x^n e^{-ax^b} = \frac{1}{b} a^{-\left(\frac{n+1}{b}\right)} \Gamma\left(\frac{n+1}{b}\right), \tag{F.8}$$

and knowing that $\Gamma(n) = (n-1)!$, equation (F.7) can be rewritten as

$$\mathbb{E}\{Z\} = \frac{1}{\Gamma(k)\theta^k}\theta^{k-1}\Gamma(k-1) = \frac{1}{\theta(k-1)}, \tag{F.9}$$

which concludes the proof. ■

Proof of the approximate MSE, η_{ik}^{prop}

For the proof of the approximate MSE, we expand it as

$$\eta_{ik}^{\text{prop}} = \frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}}\|^2\right] + \frac{1}{M}\mathbb{E}\left[\|\mathbf{g}_{iik}\|^2\right] - \frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_{iik}^{\text{prop}})^H \mathbf{g}_{iik}\right]\right], \tag{F.10}$$

and find these three expectations.

From (4.16), the first expectation can be written as

$$\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}}\|^2\right] = M\beta_{iik}^2 \mathbb{E}\left\{\frac{\|\mathbf{z}_{ik}\|^2}{[\|\mathbf{z}_{ik}\|^2]^2}\right\} = M\beta_{iik}^2 \mathbb{E}\left\{\frac{1}{\|\mathbf{z}_{ik}\|^2}\right\}. \tag{F.11}$$

From Lemma 11 we know that $\|\mathbf{z}_{ik}\|^2 \sim \Gamma(MP, \zeta_{ik})$. Then, applying Lemma 12 to (F.11) we figure out that $\mathbb{E}\{1/\|\mathbf{z}_{ik}\|^2\} = 1/\zeta_{ik}(M-1)$ and consequently, the first expectation term is defined as

$$\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}}\|^2\right] = \frac{M\beta_{iik}^2}{\zeta_{ik}(M-1)}. \tag{F.12}$$

The second expectation term is defined as

$$\frac{1}{M}\mathbb{E}\left[\|\mathbf{g}_{iik}\|^2\right] = \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\left[\|g_{iikm}\|^2\right] = \beta_{iik}. \tag{F.13}$$

Finally, in order to find the expected value of the third term, first we use (4.4) and (4.16) to rewrite it as

$$-2\beta_{iik}\mathbb{E}\left\{\Re\left[\frac{\mathbf{z}_{ik}^H\mathbf{g}_{iik}}{\|\mathbf{z}_{ik}\|^2}\right]\right\} = -2\beta_{iik}\left\{\mathbb{E}\left[\Re\left[\frac{\sum_{l=1}^{L}\mathbf{g}_{ilk}^H\mathbf{g}_{iik}}{\|\mathbf{z}_{ik}\|^2}\right]\right] + \mathbb{E}\left[\Re\left[\frac{\mathbf{w}_{ik}^H\mathbf{g}_{iik}}{\|\mathbf{z}_{ik}\|^2}\right]\right]\right\}, \tag{F.14}$$

where $\mathbf{w}_{ik} = \mathbf{N}_i\phi_k/\sqrt{q}N \sim \mathcal{CN}(\mathbf{0}_M, \frac{1}{qN}\mathbf{I}_M)$.

In order to avoid the numerical issues mentioned earlier in this work and find a simpler and more tractable equation for the MSE of the proposed channel estimator, we find approximations to the two ratios of random variables in (F.14). It is possible to approximate the moments of a function $g(X,Y)$ using Taylor series expansions, provided g is sufficiently differentiable and that the moments of X and Y are finite. Therefore, applying Lemma 9 separately to each one of the terms in the second and third lines of (F.14) we are able to find an approximation to the third expectation, which is defined as

$$\begin{aligned}-\frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_{iik}^{\text{prop}})^H\mathbf{g}_{iik}\right]\right] &\approx -2\beta_{iik}\left\{\left[\frac{\beta_{iik}}{\zeta_{ik}}\left(1-\frac{\sum_{l=1}^{L}\beta_{ilk}}{M\zeta_{ik}}+\frac{1}{M}\right)\right] + \left[\frac{-\beta_{iik}}{M\zeta_{ik}^2 qN}\right]\right\} \\ &= -\frac{2\beta_{iik}^2}{\zeta_{ik}}.\end{aligned} \tag{F.15}$$

The approximation found for (F.14) can also be elegantly found by applying Lemma 13, which is presented next, to it. After finding the three expectations, (F.12), (F.13) and (F.15), by substituting them back in the expansion of η_{ik}^{prop}, we complete the proof.

Lemma 13. *If $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_x^2\boldsymbol{I}_M)$ and $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_y^2\boldsymbol{I}_M)$ are independent, then*

$$\mathbb{E}\left\{\Re\left[\frac{(\boldsymbol{x}+\boldsymbol{y})^H\boldsymbol{x}}{\|\boldsymbol{x}+\boldsymbol{y}\|^2}\right]\right\} \approx \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}. \tag{F.16}$$

Proof. This can be proved by using an trick presented in [159] for normal random variables, which, as shown by the results, seems to be possible to be applied to complex normal random variables. We first write $\Sigma_{\mathbf{x}} = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ and $\Sigma_{\mathbf{y}} = \mathbb{E}[\mathbf{y}\mathbf{y}^H]$ as the covariance matrix of $\mathbf{x}$ and $\mathbf{y}$ respectively. Let $\mathbf{z} = \mathbf{x}+\mathbf{y}$, and $\mathbf{S} = \Sigma_{\mathbf{x}}(\Sigma_{\mathbf{x}}+\Sigma_{\mathbf{y}})^{-1}$. Therefore, we have

$$\mathbb{E}[(\mathbf{x}-\mathbf{S}\mathbf{z})\,\mathbf{z}^H] = 0, \tag{F.17}$$

so that $(\mathbf{x}-\mathbf{S}\mathbf{z})$ and $\mathbf{z}$ are uncorrelated and jointly normal, and therefore independent

$$\mathbb{E}[\mathbf{x}-\mathbf{S}\mathbf{z}]\,\mathbb{E}[\mathbf{z}]^H = 0. \tag{F.18}$$

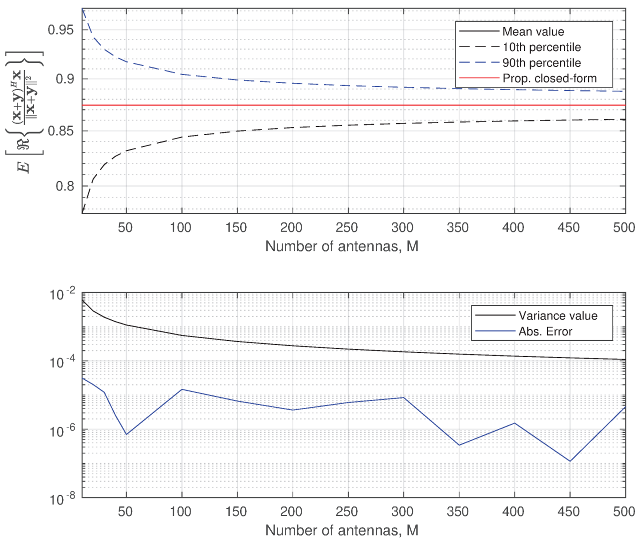

Figure F.1 – Comparison of the Monte-Carlo simulated expectation and the closed-form expression given in (F.16).

Consequently,

$$\mathbb{E}\left[\frac{\mathbf{z}^H\mathbf{x}}{\|\mathbf{z}\|^2}\right] = \mathbb{E}\left[\frac{\mathbf{z}^H(\mathbf{x}-\mathbf{S}\mathbf{z})}{\|\mathbf{z}\|^2}\right] + \mathbb{E}\left[\frac{\mathbf{z}^H\mathbf{S}\mathbf{z}}{\|\mathbf{z}\|^2}\right]$$
$$= \mathbb{E}\left[\frac{\mathbf{z}^H}{\|\mathbf{z}\|^2}\right]\mathbb{E}\left[\mathbf{x}-\mathbf{S}\mathbf{z}\right] + \mathbb{E}\left[\frac{\mathbf{z}^H\mathbf{S}\mathbf{z}}{\|\mathbf{z}\|^2}\right] = \mathbb{E}\left[\frac{\mathbf{z}^H\mathbf{S}\mathbf{z}}{\|\mathbf{z}\|^2}\right] = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{y}}^2}. \tag{F.19}$$

The result follows from the fact that $\mathbf{S} = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{y}}^2}\mathbf{I}_M$.

In Figure F.1 we present some simulation results showing the comparison of the Monte-Carlo simulated expectation and the closed-form expression given in (F.16). The figure at the top shows the simulated expectation compared to the proposed closed

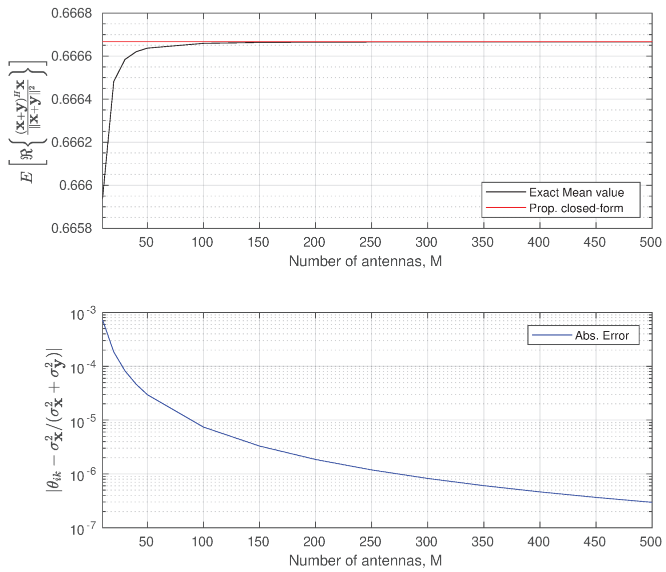

Figure F.2 – Comparison of the exact expectation, θ_{ik}, and the closed-form expression given in (F.16).

form expression along with the the 10% and 90% percentiles of the ratio $\Re\left[\frac{(\mathbf{x}+\mathbf{y})^H\mathbf{x}}{\|\mathbf{x}+\mathbf{y}\|^2}\right]$. The figure at the bottom shows the variance of the ratio and the error between its mean and $\frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}$. In Figure F.2 we compare the exact mean value, θ_{ik}, with the closed-form expression given in (F.16). The figure at the bottom shows the absolute error, and as can be noticed, the error is less than 3×10^{-5} for $M = 50$. As can be seen in figures F.1 and F.2, the closed-form expression approaches the simulated and exact values as M increases. Therefore, these results show that the proposed closed form expression is a good approximation for the exact mean value, θ_{ik}, getting more accurate as M increases.

$\blacksquare$

Appendix G – Proof of the MSE between $\hat{\mathbf{g}}_{iik}^{\text{prop}}$ and $\hat{\mathbf{g}}_{iik}^{\text{MMSE}}$ defined in Eq. (4.20).

Here we present proof for (4.20). First, we expand the normalized Euclidean distance between $\hat{\mathbf{g}}_{iik}^{\text{prop}}$ and $\hat{\mathbf{g}}_{iik}^{\text{MMSE}}$ as

$$\begin{aligned}\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}}\|^2\right] + \frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_{iik}^{\text{MMSE}}\|^2] \\ -\frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_{iik}^{\text{prop}})^H\hat{\mathbf{g}}_{iik}^{\text{MMSE}}\right]\right].\end{aligned} \tag{G.1}$$

Then we compute these three different expectations. The first one is given by (F.12), $\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}}\|^2\right] = M\beta_{iik}^2/\zeta_{ik}(M-1)$. Next, by recalling that $\hat{\mathbf{g}}_{iik}^{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M)$ we have that $\frac{1}{M}\mathbb{E}[\|\hat{\mathbf{g}}_{iik}^{\text{MMSE}}\|^2] = \beta_{iik}^2/\zeta_{ik}$. For the last expectation term, using (4.8) and (4.16), we can write it as

$$\begin{aligned}&-\frac{2}{M}\mathbb{E}\left[\Re\left[(\hat{\mathbf{g}}_{iik}^{\text{prop}})^H\hat{\mathbf{g}}_{iik}^{\text{MMSE}}\right]\right] \\ &= -\frac{2\beta_{iik}^2}{\zeta_{ik}}\mathbb{E}\left\{\Re\left[\frac{\|\mathbf{z}_{ik}\|^2}{\|\mathbf{z}_{ik}\|^2}\right]\right\} \\ &= -\frac{2\beta_{iik}^2}{\zeta_{ik}}.\end{aligned} \tag{G.2}$$

Finally, by substituting these results back into the expansion, we arrive at (4.20).

Appendix H – Proof of the covariance matrix of the columns of $\hat{\mathbf{G}}_{iik}^{\text{prop}}$ defined in Eq. (6.19).

For the proof of the covariance matrix of each one of the columns of the proposed estimator we first need to present the following Lemma.

Lemma 14. *If $X \sim \Gamma(k_1, \theta)$ and $Y \sim \Gamma(k_2, \theta)$ are independent random variables, then $U = \frac{X}{X+Y} \sim \beta(k_1, k_2)$, $V = X + Y \sim \Gamma(k_1 + k_2, \theta)$ are also independent random variables.*

Proof. To prove this, first we write the joint pdf of X and Y as

$$f_{XY}(x, y) = \frac{\theta^{k_1+k_2}}{\Gamma(k_1)\Gamma(k_2)} x^{k_1-1}\ y^{k_2-1}\ e^{-\theta(x+y)}. \tag{H.1}$$

Next, taking the Jacobian determinant of the transformation $X = UV$ and $Y = V(1 - U)$ results in

$$|J| = \left|\frac{\partial(x, y)}{\partial(u, v)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v. \tag{H.2}$$

Therefore, the joint distribution of U and V has the pdf given by

$$\begin{aligned} f_{UV}(u, v) &= f_{XY}(x, y) \left|\frac{\partial(x, y)}{\partial(u, v)}\right| \\ &= \frac{\theta^{k_1+k_2}}{\Gamma(k_1)\Gamma(k_2)} (uv)^{k_1-1}\ [v(1-u)]^{k_2-1}\ e^{-\theta(v)} v \\ &= \left\{\frac{\theta^{(k_1+k_2)}}{\Gamma(k_1+k_2)} v^{k_1+k_2-1} e^{-\theta v}\right\} \left\{\frac{u^{k_1-1}(1-u)^{k_2-1}}{\beta(k_1, k_2)}\right\}, \end{aligned} \tag{H.3}$$

and hence, as is apparent from the two terms separated by the curly brackets, U and V are independent random variables with $V \sim \Gamma(k_1 + k_2, \theta)$ and $U \sim \beta(k_1, k_2)$. ∎

Proof of the covariance matrix

With the purpose of helping this and the next proofs, the LS estimator can be expressed in the following way

$$\mathbf{Z}_{ik} = \begin{bmatrix} z_{ik1}^{(0)} & z_{ik1}^{(1)} & \cdots & z_{ik1}^{(\Psi_P-1)} \\ z_{ik2}^{(0)} & z_{ik2}^{(1)} & \cdots & z_{ik2}^{(\Psi_P-1)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{ikM}^{(0)} & z_{ikM}^{(1)} & \cdots & z_{ikM}^{(\Psi_P-1)} \end{bmatrix}, \tag{H.4}$$

where $z_{ikm}^{(p)} = \sum_{l=1}^{L} g_{ilkm}^{(p)} + w_{ikm}^{(p)} \sim \mathcal{CN}(0, \zeta_{ik})\ \forall p$.

It is important to observe that the Trace of $\mathbf{Z}_{ik}^{H}\mathbf{Z}_{ik}$, *i.e.*, $\text{Tr}(\|\mathbf{Z}_{ik}\|^2)$, results in a positive real-valued scalar random variable, $\|\mathbf{Z}_{ik}\|_F^2 = \sum_{m=1}^{M}\sum_{p=0}^{\Psi_P-1} |z_{ikm}^{(p)}|^2$. Additionally, by using Lemmas 9 and 14 we find that $\text{Tr}(\|\mathbf{Z}_{ik}\|^2) \sim \Gamma(M\Psi_P, \zeta_{ik})$.

We start the proof of (6.19) by re-writing the estimated channel matrix (6.18) as

$$\begin{aligned}\hat{\mathbf{G}}_{iik}^{\text{prop}} &= a \left[\begin{array}{cccc} \mathbf{z}_{ik}^{(0)} & \mathbf{z}_{ik}^{(1)} & \cdots & \mathbf{z}_{ik}^{(\Psi_P-1)} \end{array}\right] \\ &= a \begin{bmatrix} z_{ik1}^{(0)} & z_{ik1}^{(1)} & \cdots & z_{ik1}^{(\Psi_P-1)} \\ z_{ik2}^{(0)} & z_{ik2}^{(1)} & \cdots & z_{ik2}^{(\Psi_P-1)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{ikM}^{(0)} & z_{ikM}^{(1)} & \cdots & z_{ikM}^{(\Psi_P-1)} \end{bmatrix}.\end{aligned} \tag{H.5}$$

where $a = M\Psi_P \frac{\beta_{iik}}{\|\mathbf{Z}_{ik}\|_F^2}$. The covariance matrix of each one of the columns of $\hat{\mathbf{G}}_{iik}^{\text{prop}}$ can be expressed as

$$\begin{aligned}&\mathbb{E}\left[\mathbf{z}_{ik}^{(p)}\left(\mathbf{z}_{ik}^{(p)}\right)^H\right] \\ &= (M\Psi_P\beta_{iik})^2 \begin{bmatrix} \mathbb{E}\left\{\frac{|z_{ik1}^{(p)}|^2}{\left(\|\mathbf{Z}_{ik}\|_F^2\right)^2}\right\} & \cdots & \mathbb{E}\left\{\frac{z_{ik1}^{(p)}(z_{ikM}^{(p)})^*}{\left(\|\mathbf{Z}_{ik}\|_F^2\right)^2}\right\} \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left\{\frac{z_{ikM}^{(p)}(z_{ik1}^{(p)})^*}{\left(\|\mathbf{Z}_{ik}\|_F^2\right)^2}\right\} & \cdots & \mathbb{E}\left\{\frac{|z_{ikM}^{(\Psi_P-1)}|^2}{\left(\|\mathbf{Z}_{ik}\|_F^2\right)^2}\right\} \end{bmatrix}.\end{aligned} \tag{H.6}$$

Next, by applying Lemmas 14 and 12 to the elements on the main diagonal and Lemma 9 to the elements on the off-diagonal of (H.7) we find

$$\begin{aligned}&\mathbb{E}\left[\mathbf{z}_{ik}^{(p)}\left(\mathbf{z}_{ik}^{(p)}\right)^H\right] \\ &= (M\Psi_P\beta_{iik})^2 \begin{bmatrix} \frac{1}{M\Psi_P(M\Psi_P-1)\zeta_{ik}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{M\Psi_P(M\Psi_P-1)\zeta_{ik}} \end{bmatrix},\end{aligned} \tag{H.7}$$

which concludes the proof.

The expectation of the elements on the off-diagonal of (H.7) can also be found by using the symmetry argument. First, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}\left[\left|\frac{z_{ikm}^{(p)}}{\|\mathbf{Z}_{ik}\|_F^2}\frac{(z_{ikn}^{(p)})^*}{\|\mathbf{Z}_{ik}\|_F^2}\right|\right] &\leq \sqrt{\mathbb{E}\left[\left|\frac{z_{ikm}^{(p)}}{\|\mathbf{Z}_{ik}\|_F^2}\right|^2\right]\mathbb{E}\left[\left|\frac{z_{ikn}^{(p)}}{\|\mathbf{Z}_{ik}\|_F^2}\right|^2\right]} \\ &= \mathbb{E}\left[\left|\frac{z_{ikm}^{(p)}}{\|\mathbf{Z}_{ik}\|_F^2}\right|^2\right] < \infty, m \neq n,\end{aligned} \tag{H.8}$$

hence, $\mathbb{E}\left[\frac{z_{ikm}^{(p)}(z_{ikn}^{(p)})^*}{\left(\|\mathbf{Z}_{ik}\|_F^2\right)^2}\right]$ exists and is finite. Therefore and because the joint distribution of the pair $\left(-\frac{z_{ikm}^{(p)}}{\|\mathbf{Z}_{ik}\|_F^2}, \frac{(z_{ikn}^{(p)})^*}{\|\mathbf{Z}_{ik}\|_F^2}\right)$ is the same as that of $\left(\frac{z_{ikm}^{(p)}}{\|\mathbf{Z}_{ik}\|_F^2}, \frac{(z_{ikn}^{(p)})^*}{\|\mathbf{Z}_{ik}\|_F^2}\right)$, we conclude that

$$\mathbb{E}\left[\frac{z_{ikm}^{(p)}(z_{ikn}^{(p)})^*}{\left(\|\mathbf{Z}_{ik}\|_F^2\right)^2}\right] = \mathbb{E}\left[-\frac{z_{ikm}^{(p)}(z_{ikn}^{(p)})^*}{\left(\|\mathbf{Z}_{ik}\|_F^2\right)^2}\right] = 0. \tag{H.9}$$

Appendix I – Proof of the approximate MSE defined in Eq. (6.20).

For the proof of the approximate MSE defined in Eq. (6.20), we first expand η_{ik}^{prop} as

$$\eta_{ik}^{\text{prop}} = \frac{1}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\hat{\mathbf{G}}_{iik}^{\text{prop}}(\hat{\mathbf{G}}_{iik}^{\text{prop}})^H\right]\right\}$$
$$+\frac{1}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\mathbf{G}_{iik}(\mathbf{G}_{iik})^H\right]\right\} - \frac{2}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\Re\left\{\hat{\mathbf{G}}_{iik}^{\text{prop}}(\mathbf{G}_{iik})^H\right\}\right]\right\}, \quad \text{(I.1)}$$

and find these three expected values.

From (6.17), the first expectation can be written as

$$\begin{aligned}
&\frac{1}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\hat{\mathbf{G}}_{iik}^{\text{prop}}(\hat{\mathbf{G}}_{iik}^{\text{prop}})^H\right]\right\} \\
&= M\Psi_P\beta_{iik}^2\mathbb{E}\left\{\frac{\text{Tr}\left[\mathbf{Z}_{ik}\mathbf{Z}_{ik}^H\right]}{[\|\mathbf{Z}_{ik}\|_F^2]^2}\right\} \\
&= M\Psi_P\beta_{iik}^2\mathbb{E}\left\{\frac{1}{\|\mathbf{Z}_{ik}\|_F^2}\right\}.
\end{aligned} \quad \text{(I.2)}$$

From Lemma 11 we know that $\|\mathbf{Z}_{ik}\|_F^2 \sim \Gamma(M\Psi_P, \zeta_{ik})$. Then, applying Lemma 12 to (I.2) we find that $\mathbb{E}\{1/\|\mathbf{Z}_{ik}\|_F^2\} = 1/\zeta_{ik}(M\Psi_P - 1)$ and consequently, the first expectation term is defined as

$$\frac{1}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\hat{\mathbf{G}}_{iik}^{\text{prop}}(\hat{\mathbf{G}}_{iik}^{\text{prop}})^H\right]\right\} = \frac{M\Psi_P\beta_{iik}^2}{\zeta_{ik}(M\Psi_P - 1)}. \quad \text{(I.3)}$$

The second expectation term is defined as

$$\frac{1}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\mathbf{G}_{iik}(\mathbf{G}_{iik})^H\right]\right\} = \sum_{p=0}^{\Psi_P-1}\sum_{m=1}^{M}\mathbb{E}\{|g_{iikm}^{(p)}|^2\} = \beta_{iik}. \quad \text{(I.4)}$$

Finally, in order to find the expected value of the third term, first we use (6.6) and (6.17) to rewrite it as

$$\begin{aligned}
&-2\beta_{iik}\mathbb{E}\left\{\Re\left[\frac{\text{Tr}\left(\mathbf{Z}_{ik}\mathbf{G}_{iik}^H\right)}{\|\mathbf{Z}_{ik}\|_F^2}\right]\right\} \\
&= -2\beta_{iik}\left\{\mathbb{E}\left[\Re\left[\frac{\sum_{m=1}^{M}\sum_{p=0}^{\Psi_P-1}\sum_{l=1}^{L} g_{ilkm}^{(p)} g_{iikm}^{\star(p)}}{\sum_{m=1}^{M}\sum_{p=0}^{\Psi_P-1}|z_{ikm}^{(p)}|^2}\right]\right]\right. \\
&\quad \left. + \mathbb{E}\left[\Re\left[\frac{\sum_{m=1}^{M}\sum_{p=0}^{\Psi_P-1} w_{ikm}^{(p)} g_{iikm}^{\star(p)}}{\sum_{m=1}^{M}\sum_{p=0}^{\Psi_P-1}|z_{ikm}^{(p)}|^2}\right]\right]\right\}.
\end{aligned} \quad \text{(I.5)}$$

The two ratios of random variables in (I.5) do not have a well-defined expectation, *i.e.* there is no closed formula for them. However, it is possible to approximate the moments of a function $g(X, Y)$ using Taylor series expansions, provided g is sufficiently differentiable and that the moments of X and Y are finite. Therefore, applying Lemma 9 to both terms in (I.5) we find an approximation to the third expectation

$$-\frac{2}{M\Psi_P}\mathbb{E}\left\{\mathrm{Tr}\left[\Re\left\{\hat{\mathbf{G}}_{iik}^{\mathrm{prop}}(\mathbf{G}_{iik})^H\right\}\right]\right\} \approx -\frac{2\beta_{iik}^2}{\zeta_{ik}}. \tag{I.6}$$

We complete the proof by substituting (I.3), (I.4) and (I.6) back in the expansion of $\eta_{ik}^{\mathrm{prop}}$ defined by (I.1).

Appendix J – Proof of the MSE between $\hat{\mathbf{G}}_{iik}^{\text{prop}}$ and $\hat{\mathbf{G}}_{iik}^{\text{MMSE}}$ defined in (6.21).

In this Appendix, we present proof for (6.21). First, we expand the Euclidean distance between $\hat{\mathbf{G}}_{iik}^{\text{prop}}$ and $\hat{\mathbf{G}}_{iik}^{\text{MMSE}}$ as

$$\begin{aligned}\frac{1}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\hat{\mathbf{G}}_{iik}^{\text{prop}}(\hat{\mathbf{G}}_{iik}^{\text{prop}})^H\right]\right\} + \frac{1}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\hat{\mathbf{G}}_{iik}^{\text{MMSE}}(\hat{\mathbf{G}}_{iik}^{\text{MMSE}})^H\right]\right\} \\ -\frac{2}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\Re\left\{\hat{\mathbf{G}}_{iik}^{\text{prop}}(\hat{\mathbf{G}}_{iik}^{\text{MMSE}})^H\right\}\right]\right\}.\end{aligned} \tag{J.1}$$

Then we compute these three different expectations. The first one is given by (I.3), $\frac{1}{M\Psi_P}\mathbb{E}\{\text{Tr}[\hat{\mathbf{G}}_{iik}^{\text{prop}}(\hat{\mathbf{G}}_{iik}^{\text{prop}})^H]\} = M\Psi_P\beta_{iik}^2/\zeta_{ik}(M\Psi_P - 1)$. Next, by recalling that $\hat{\mathbf{G}}_{iik}^{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M)$ we have that $\frac{1}{M\Psi_P}\mathbb{E}\{\text{Tr}[\hat{\mathbf{G}}_{iik}^{\text{MMSE}}(\hat{\mathbf{G}}_{iik}^{\text{MMSE}})^H]\} = \beta_{iik}^2/\zeta_{ik}$. For the last expectation term, using (6.10) and (6.17), we can write it as

$$-\frac{2}{M\Psi_P}\mathbb{E}\left\{\text{Tr}\left[\Re\left\{\hat{\mathbf{G}}_{iik}^{\text{prop}}(\hat{\mathbf{G}}_{iik}^{\text{MMSE}})^H\right\}\right]\right\} = -\frac{2\beta_{iik}^2}{\zeta_{ik}}\mathbb{E}\left\{\frac{\text{Tr}(\mathbf{Z}_{ik}\mathbf{Z}_{ik}^H)}{\text{Tr}(\mathbf{Z}_{ik}^H\mathbf{Z}_{ik})}\right\} = -\frac{2\beta_{iik}^2}{\zeta_{ik}}. \tag{J.2}$$

Finally, by substituting these results back into the expansion in (J.1), we arrive at (6.21).