

Department of Information and Communications Engineering

Resource optimization for massive MIMO systems

Kameswara Atchutaram Kocharalakota



Resource optimization for massive MIMO systems

Kameswara Atchutaram Kocharlakota

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall 1501 Sklidowska-Curie of the school on 20 December 2024 at 12:00.

Aalto University
School of Electrical Engineering
Department of Information and Communications Engineering

Supervising professor

Professor Sergiy A. Vorobyov, Aalto University, Finland

Thesis advisor

Professor Sergiy A. Vorobyov, Aalto University, Finland

Preliminary examiners

Professor Gayan Amarasuriya Aruma Baduge, Southern Illinois University, IL, USA

Professor Sundeep Prabhakar Chepuri, Indian Institute of Science, India

Opponents

Professor Gayan Amarasuriya Aruma Baduge, Southern Illinois University, IL, USA

Professor Angelo Coluccia, University of Salento, Lecce, Italy

Aalto University publication series

DOCTORAL THESES 278/2024

© 2024 Kameswara Atchutaram Kocharlakota

ISBN 978-952-64-2198-8 (printed)

ISBN 978-952-64-2199-5 (pdf)

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

<http://urn.fi/URN:ISBN:978-952-64-2199-5>

Images: Image by Erich Westendarp from Pixabay (cover photo)

Unigrafia Oy

Helsinki 2024

Finland



Printed matter
4041-0619

Author

Kameswara Atchutaram Kocharalakota

Name of the doctoral thesis

Resource optimization for massive MIMO systems

Publisher School of Electrical Engineering**Unit** Department of Information and Communications Engineering**Series** Aalto University publication series DOCTORAL THESES 278/2024**Field of research** Signal Processing Technology**Manuscript submitted** 17 June 2024**Date of the defence** 20 December 2024**Permission for public defence granted (date)** 13 September 2024**Language** English **Monograph** **Article thesis** **Essay thesis****Abstract**

The thesis delves into the intricacies of resource optimization in both massive Multiple-Input Multiple-Output (MIMO) and cell-free massive MIMO (CFmMIMO) systems, which are pivotal for the advancement of 5G and beyond wireless networks. The research primarily addresses the challenges of pilot contamination and power allocation, which significantly impact the spectral efficiency (SE) and overall performance of these systems.

Initially, the thesis explores the massive MIMO systems, focusing on the impact of pilot overhead and the accuracy of channel estimation on the SE. Closed-form expressions for the uplink (UL) and downlink (DL) SEs under conditions of imperfect channel state information (CSI) are derived. These expressions are crucial in understanding the trade-offs involved in pilot resource allocation, emphasizing that efficient pilot management is essential for maintaining high system performance. The analysis provides the closed-form expressions as vital tools for selecting optimal pilot overhead parameters.

In the latter part, the thesis shifts focus to CFmMIMO systems, which distribute antennas across a large area to provide uniform coverage and enhance the performance of cell-edge users. Here, the primary challenge addressed is the downlink power control. Traditional methods for power control are computationally intensive and often inadequate for the centralized nature of CFmMIMO systems. To overcome these limitations, the research introduces advanced deep learning techniques, specifically Attention Neural Networks (ANN) and Pilot Contamination-Aware Power Control (PAPC) transformer neural network, for power control. These models leverage the capabilities of masked multi-head attention networks, enabling efficient power allocation even in the presence of pilot contamination.

The ANN-based approach initially transforms the constrained optimization problem into an unconstrained one, optimized through unsupervised learning. Subsequently, PAPC further refines this approach by incorporating additional architectural enhancements, such as pre-processing and post-processing stages, which improve performance and scalability while reducing computational complexity. Extensive simulations validate the effectiveness of these proposed solutions, demonstrating their potential to significantly reduce the computational complexity while providing state-of-the-art performance in CFmMIMO systems.

In conclusion, this thesis makes significant contributions to the field of wireless communications by providing innovative solutions and comprehensive analytical tools for resource optimization in massive MIMO and CFmMIMO systems. The findings and methodologies presented are expected to pave the way for more efficient and reliable next-generation wireless communication technologies, addressing critical challenges in pilot resource allocation and power control.

Keywords Massive MIMO, Cell-Free Massive MIMO, Resource Allocation, Pilot Contamination, Power Allocation, Deep Learning in Wireless Communication, GPT, BERT

ISBN (printed) 978-952-64-2198-8**ISBN (pdf)** 978-952-64-2199-5**ISSN (printed)** 1799-4934**ISSN (pdf)** 1799-4942**Location of publisher** Helsinki**Location of printing** Helsinki**Year** 2024**Pages** 112**urn** <http://urn.fi/URN:ISBN:978-952-64-2199-5>

Preface

This thesis research was conducted under the guidance of Prof. Sergiy A. Vorobyov at the Department of Information & Communications Engineering, Aalto University. His steadfast support and understanding were invaluable during the challenges posed by the COVID-19 pandemic, ensuring the continuity of this work despite unprecedented global disruptions.

Special thanks to Prof. Gayan Amarasuriya Aruma Baduge and Prof. Angelo Coluccia for serving as opponents for my defense. I am equally thankful to my pre-examiners, Prof. Gayan Amarasuriya Aruma Baduge and Prof. Sundeep Prabhakar Chepuri, for their meticulous review and constructive criticism.

I am grateful to Prof. Visa Koivunen and Prof. Esa Ollila for their insightful courses and feedback during seminars, and to Prof. Robert W. Heath Jr. for his guidance and valuable comments on our co-authored publications.

I am thankful to my colleagues at Aalto University for making this journey enjoyable, especially Dr. Karthik Upadhyia for his support in our publications and for suggesting this place to begin my studies, and to Palshin Andrei for his assistance with simulations.

I am grateful to my friends outside academia for their steady support and encouragement.

I am indebted to my brother and my entire family for their patience and enduring my prolonged absence during my research. I wish to acknowledge my late parents, whose love, values, and teachings remain a lasting foundation of strength, inspiration, and accomplishment in my life.

Lastly, and most importantly, I owe profound gratitude to my wife for her love, understanding, and sacrifices which played an essential role in helping me complete this work. Our shared strength and mutual support during this period have been the foundation of my perseverance and the completion of this work.

Espoo, November 26, 2024,

Kameswara Atchutaram Kocharlakota

Contents

| | |
|--|------------|
| Preface | i |
| Contents | iii |
| List of Publications | v |
| Author's Contribution | vii |
| Abbreviations | ix |
| Symbols | xi |
| 1. Introduction | 1 |
| 1.1 Background and Motivation | 1 |
| 1.2 Objectives of the Research | 2 |
| 1.3 Contributions of the Thesis | 2 |
| 1.4 Structure of the Thesis | 4 |
| 2. Massive MIMO and Cell-Free Massive MIMO Systems | 5 |
| 2.1 Introduction | 5 |
| 2.2 Massive MIMO Systems | 6 |
| 2.2.1 System Description | 6 |
| 2.2.2 TDD for Massive MIMO | 7 |
| 2.2.3 System Model | 7 |
| 2.2.4 Pilot Contamination | 8 |
| 2.2.5 Channel Estimation | 9 |
| 2.2.6 Spectral Efficiency Analysis for Uplink and Downlink | 9 |
| 2.2.7 Imperfect CSI: Challenges in Pilot Resource Allocation | 10 |
| 2.3 Cell-Free Massive MIMO Systems | 12 |
| 2.3.1 System Description | 12 |
| 2.3.2 System Model | 13 |
| 2.3.3 Power Allocation in Cell-Free Massive MIMO | 14 |
| 2.4 Conclusion | 15 |

| | |
|---|-----------|
| 3. Performance Analysis of Massive MIMO Systems with Pilot Resource Allocation | 17 |
| 3.1 Introduction | 17 |
| 3.2 Channel Estimation | 17 |
| 3.2.1 LMMSE-type Channel Estimation | 18 |
| 3.2.2 Element-wise LMMSE-type Channel Estimation | 18 |
| 3.2.3 Covariance Matrix Estimation | 18 |
| 3.3 Spectral Efficiency Metrics | 20 |
| 3.3.1 Uplink Spectral Efficiency | 20 |
| 3.3.2 Downlink Spectral Efficiency | 22 |
| 3.4 Contributions | 22 |
| 3.5 Average Spectral Efficiency and Impact of Pilot Overhead | 23 |
| | |
| 4. Transformer Neural Network for Downlink Power Control in Cell-Free Massive MIMO Systems | 27 |
| 4.1 Introduction | 27 |
| 4.2 Unsupervised Learning Approach for Downlink Power Control | 27 |
| 4.3 Literature Review on Deep learning Based Solutions | 29 |
| 4.4 Contributions | 29 |
| 4.4.1 Why Transformer-Based Models for CFmMIMO? | 30 |
| 4.4.2 Overview of ANN Model | 30 |
| 4.4.3 Transition to the PAPC Model (Publication IV) | 30 |
| 4.5 Results | 33 |
| 4.5.1 Simulation Setup | 33 |
| 4.5.2 Results Summary | 34 |
| | |
| 5. Conclusion and Future Work | 39 |
| 5.1 Conclusion | 39 |
| 5.2 Future Work | 40 |
| | |
| References | 43 |
| | |
| Errata | 47 |
| | |
| Publications | 49 |

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

- I** A. K. Kocharlakota, K. Upadhyaya and S. A. Vorobyov. On Achievable Rates for Massive MIMO System with Imperfect Channel Covariance Information. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, pp. 4504-4508, 2019.
- II** A. K. Kocharlakota, K. Upadhyaya and S. A. Vorobyov. Impact of Pilot Overhead and Channel Estimation on the Performance of Massive MIMO. *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8242-8255, Dec 2021.
- III** A. K. Kocharlakota, S. A. Vorobyov and R. W. Heath. Attention Neural Network for Downlink Cell-Free Massive MIMO Power Control. In *Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp. 738-743, 2022.
- IV** A. K. Kocharlakota, S. A. Vorobyov and R. W. Heath. Pilot Contamination Aware Transformer Neural Network for Downlink Power Control in Cell-Free Massive MIMO Networks. *Submitted to IEEE Trans. Wireless Commun.*, Jun 2024.

Author's Contribution

Publication I: "On Achievable Rates for Massive MIMO System with Imperfect Channel Covariance Information"

The author developed the core analysis, performed the numerical experiments, interpreted the results, and wrote the article, integrating feedback from the co-authors.

Publication II: "Impact of Pilot Overhead and Channel Estimation on the Performance of Massive MIMO"

The author developed the core analysis, performed the numerical experiments, interpreted the results, and wrote the article, integrating feedback from the co-authors.

Publication III: "Attention Neural Network for Downlink Cell-Free Massive MIMO Power Control"

The author developed the idea, designed and implemented the deep learning algorithms, conducted all numerical experiments, analyzed the results, and wrote the article, incorporating feedback from the co-authors.

Publication IV: "Pilot Contamination Aware Transformer Neural Network for Downlink Power Control in Cell-Free Massive MIMO Networks"

The author developed the idea, designed and implemented the deep learning algorithms, conducted all numerical experiments, analyzed the results, and wrote the article, incorporating feedback from the co-authors.

Abbreviations

ANN Attention Neural Network

AP Access Point

APG Accelerated Proximal Gradient

AWGN Additive White Gaussian Noise

BS Base Station

CFmMIMO Cell-Free massive Multiple-Input Multiple-Output

CPU Central Processing Unit

CSI Channel State Information

DL Downlink

DNN Deep Neural Network

EPA Equal Power Allocation

FCN Fully Connected Network

GPU Graphics Processing Unit

LMMSE Linear Minimum Mean Square Error

LS Least Squares

LSFC Large-Scale Fading Coefficient

MIMO Multiple-Input Multiple-Output

MF Matched Filter

MMHAN Masked Multi Head Attention Network

MMSE Minimum Mean Square Error

Abbreviations

MRC Maximum Ratio Combining

PAPC Pilot contamination-Aware Power Control

SCA Successive Convex Approximation

SE Spectral Efficiency

SINR Signal-to-Interference-plus-Noise Ratio

SNR Signal-to-Noise Ratio

TDD Time Division Duplex

UL Uplink

ZF Zero Forcing

Symbols

Latin Letters

A Matrix (boldface capital letter)

a Vector (boldface lowercase letter)

\mathbf{b}_{lk} Corresponding precoding vector

C_u Number of symbols for uplink communication

g_{mk} Channel between the k -th user and the m -th AP/BS

h_{jlk} Channel between user (l, k) and BS j

\mathbf{h}_{mk} Small-scale fading coefficients

I Identity matrix (of appropriate size)

K Number of users per cell

K Number of single-antenna users

L Number of cells

M Number of antennas at the base station (BS)

M Number of access points (APs)

M Matrix of all power control coefficients

N Additive white Gaussian noise (AWGN)

N Number of antennas at each AP/BS

\mathbf{R}_m Uplink received signal at the m -th AP/BS

\mathbf{R}_{jlk} Covariance matrix of h_{jlk}

\mathbf{r}_k Received signal at the k -th user

Symbols

W Combining vector for MRC or ZF

$\mathbf{x}_{l,k}$ Signal transmitted by user (l, k)

\mathbf{x}_m Beamformed downlink signal at the m -th AP/BS

Y UL received signal

$\mathbf{Z}_{p,m}$ Noise matrix

$\mathbf{z}_{d,m}$ Noise signal

Greek Letters

β_{mk} Large-scale fading coefficient (LSFC)

δ_{ij} Kronecker delta, $\delta_{ij} = 1$ if $i = j$, and 0 otherwise

μ Transmit power at each user

μ_m Vector of power control coefficients associated with the m -th AP/BS

ν_{mk} Mean square value of the estimated channel

τ Number of symbols within a coherence block

τ_c Length of the coherence block

τ_p Number of symbols for pilot transmission

ψ_k Pilot sequence transmitted by the k -th user

Notation

\mathbb{R} Set of real numbers

\mathbb{R}_+ Set of positive real numbers

\mathbb{C} Set of complex numbers

$(\cdot)^*$ Element-wise conjugate

$(\cdot)^\top$ Transpose

$(\cdot)^H$ Hermitian transpose

\odot Hadamard product (element-wise product)

$\|\cdot\|$ l_2 norm of a vector or matrix (Frobenius norm)

$\mathbb{E}\{\cdot\}$ Mathematical expectation

$\ln(\cdot)$ Natural logarithm

$\text{tr}(\cdot)$ Trace of a matrix

$\text{diag}(\mathbf{A})$ Diagonal matrix with the same diagonal elements as \mathbf{A}

$\mathcal{CN}(\mathbf{m}, \mathbf{R})$ Circularly symmetric complex Gaussian random vector with mean vector \mathbf{m} and covariance matrix \mathbf{R}

$\mathcal{W}(N, \mathbf{R})$ Wishart random matrix with N degrees of freedom and covariance matrix \mathbf{R}

1. Introduction

1.1 Background and Motivation

The evolution of wireless networks has positioned massive Multiple-Input Multiple-Output (MIMO) as a pivotal technology for both fifth-generation (5G) and forthcoming beyond 5G networks. Massive MIMO is characterized by the deployment of a significantly large number of antennas at the base station (BS) within a cell to serve multiple users concurrently. This concept offers significant benefits, including increased spectral efficiency (SE) and energy efficiency, as well as enhanced system capacity [1–5] with low computational cost [1, 6, 7]. The architecture has been shown to harness the spatial domain to its maximum potential, effectively multiplying the capacity of a wireless link.

Moving beyond the conventional cell-based system, Cell-Free massive Multiple-Input Multiple-Output (CFmMIMO) presents a promising advancement. Contrary to the traditional cellular approach, where each user is associated with a specific BS or cell, in CFmMIMO, all the antennas distributed across a large geographical area simultaneously serve all the users. This provides superior benefits, such as improved performance for cell-edge users and circumvents handover issues of cellular systems [8]. The essence of the cell-free paradigm lies in the coherent processing of signals from all antennas, which can provide uniform coverage throughout the network area, leading to a more efficient and reliable wireless communication system [8–13].

Massive MIMO and CFmMIMO systems, while being innovative and promising technologies, face a host of challenges, especially concerning resource allocation. Addressing resource allocation issues is crucial for the successful deployment and performance of these systems. Key among these challenges are the pilot and power allocation algorithms, both of which have a significant bearing on the system performance of massive MIMO. The substantial increase in the number of antennas in massive MIMO systems necessitates a corresponding escalation in pilot overhead for reliable channel estimation. However, the feasibility of this increase is constrained by the coherence bandwidth and coherence time,

leading to a finite amount of available resources (transmit symbols) for pilot transmission. This limit on pilot overhead demands a meticulous and strategic allocation to optimize system efficiency. Furthermore, power allocation presents another critical resource allocation challenge. The difficulty lies in optimally distributing power across multiple channels to minimize inter-user interference while concurrently ensuring that each user receives sufficient power for reliable communication. Power allocation is especially extremely challenging in the DL of a CFmMIMO due to the centralized nature of its architecture.

This thesis aims to address the aforementioned challenges by focusing on the optimization of resource allocation in massive MIMO and CFmMIMO systems, with particular emphasis on pilot and power allocation. The motivation behind this research focus stems from the critical impact these aspects have on the overall performance of wireless communication systems. Efficient allocation of pilot sequences is vital to accurate channel estimation and system throughput, while robust power control schemes are necessary to minimize inter-user interference and ensure reliable communication for all users.

1.2 Objectives of the Research

The core objectives guiding this research endeavor can be succinctly stated as follows:

- Initiate a detailed study on the impact of pilot overhead and the accuracy of channel estimation on the SE in massive MIMO systems. A key aim is to develop closed-form expressions for uplink (UL) and downlink (DL) SEs of massive MIMO systems with imperfect channel information. These expressions are expected to provide significant insights, thereby aiding in the creation of strategies for efficient pilot resource allocation.
- Conduct a comprehensive investigation of the DL power control issue prevalent in CFmMIMO systems. The goal is to propose an effective solution, underpinned by deep learning techniques, to address this innate challenge, hence improving the overall performance of the system.

1.3 Contributions of the Thesis

This thesis furnishes significant advancements in understanding and addressing key challenges in massive MIMO and CFmMIMO systems, as detailed in the following research publications:

- In Publication I, we delve into the theoretical aspects of the UL SE for a user in

a massive MIMO system, establishing an analytical lower bound, particularly under conditions when the channel vector and covariance matrices are not known or readily available. The derived bounds are expressed as a function of pilot overhead, providing a clear quantification of the number of samples needed for covariance estimation to meet specific SE targets. The comparison between the analytical bounds and simulated outcomes confirms the reliability of the lower bound.

- Publication II presents our derivation of analytical expressions encapsulating both the UL and DL SEs of a massive MIMO system with imperfect Channel State Information (CSI). These are explicitly formulated as functions of pilot overhead, while the massive MIMO system employs Linear Minimum Mean Square Error (LMMSE) and element-wise LMMSE channel estimation methodologies. The analysis highlights how pilot overhead affects the accuracy of covariance matrix estimation, thereby impacting the overall SE performance. A detailed theoretical discussion quantifies the SE behavior as a function of pilot overhead, capturing the trade-off between additional pilots and the improvement in channel estimation quality.
- In Publication III, we introduce a novel solution to the prevalent DL power control problem in CFmMIMO systems. Here we design a new Attention Neural Network (ANN) that harnesses the capabilities of the Masked Multi Head Attention Network (MMHAN) module, which is a building block of the popular transformers [14]. The model's robustness is proven through comprehensive numerical simulations and an analysis of computational complexity is given. The proposed network efficiently addresses pilot contamination by incorporating pilot allocation information in its design, and significantly reduces the computational cost compared to traditional optimization-based methods.
- In a significant step forward, Publication IV refines and extends the approach introduced in Publication III by developing a Pilot Contamination-Aware Power Control (PAPC) transformer neural network for DL power control in CFmMIMO systems. This work introduces crucial architectural enhancements, such as additional pre-processing and post-processing stages, and an improved objective function, resulting in enhanced performance. While PAPC retains the core principles of the ANN, these modifications not only improve performance but also enhance scalability and adaptability in CFmMIMO networks. By integrating pilot allocation information, PAPC effectively reduces the negative impact of pilot contamination, improving the quality of power control in large-scale systems. Additionally, PAPC's ability to dynamically adapt to varying numbers of users without retraining demonstrates its flexibility to varying number of users. The PAPC model effectively handles pilot contamination and demonstrates a significant reduction in computational complexity compared to traditional methods. Through extensive simulations, this study

showcases the PAPC's ability to match the performance of state-of-the-art algorithms, emphasizing its robustness and effectiveness in complex CFmMIMO environments.

1.4 Structure of the Thesis

Chapter 2 offers a comprehensive overview of massive MIMO technology, presenting its fundamental concepts and inherent challenges. It also introduces CFmMIMO systems, highlighting their distinctive advantages over traditional cell-based systems.

Chapter 3 delves into the implications of pilot overhead and the accuracy of channel estimation on the SE in massive MIMO systems. Drawing upon the insights from Publications I and II, this chapter conducts an extensive analysis of pilot resource allocation.

Chapter 4 presents an advanced approach to DL power control in CFmMIMO systems using PAPC. This chapter addresses the complexities of power control, including the non-convex nature of the problem and the inefficiencies of traditional iterative methods. It begins with describing an unsupervised learning framework where the neural network optimizes power allocation without the need for extensive labeled data. Initially, an ANN is introduced, leveraging MMHAN modules, as detailed in Publication III. Building on this foundation, the chapter introduces PAPC developed in Publication IV, which retains the core principles of the ANN but incorporates significant architectural enhancements, resulting in improved performance, and a reduction in computational complexity, as demonstrated by extensive simulations.

Chapter 5 serves as the culmination of the thesis, summarizing the key findings and contributions, and proposing directions for future research within the domain of massive MIMO and CFmMIMO resource allocation.

2. Massive MIMO and Cell-Free Massive MIMO Systems

2.1 Introduction

In the ever-evolving landscape of wireless communication, MIMO has emerged as a crucial cornerstone technology. This transformative technology, especially in its ‘massive’ variant, holds the potential to redefine the frontiers of capacity, energy efficiency, and robustness. Massive MIMO, often considered synonymous with large-scale antenna systems, involves BSs equipped with a large number of antennas that simultaneously serve multiple users on the same time-frequency resources. Therefore, this is a multi-user MIMO setup, often referred to as multi-user massive MIMO or MU-massive-MIMO. The spatial multiplexing capabilities due to the abundance of antennas at the BS has the potential to increase the system capacity by orders of magnitude, thus enabling a paradigm shift in network design. Furthermore, by virtue of its design, massive MIMO can enhance the system’s robustness against several disruptive factors, such as interference and fading [1–5, 15].

As an extension of massive MIMO technology, CFmMIMO has started to take form, introducing a notable evolution in the field of wireless communication. This technology spreads the multitude of antennas in massive MIMO systems across a large geographical spread, rather than confining them to discrete BSs. This design approach substantially enhances the robustness of the system against disruptions from shadowing and fading effects, while improving the user experience due to reduced inter-user interference. Its potential for providing uniformly good coverage to all users makes CFmMIMO an attractive prospect for beyond 5G wireless networks [8, 16, 17].

Successful implementation of massive MIMO and CFmMIMO systems involves overcoming several challenges. The benefits of massive MIMO systems are predominantly subject to the availability of perfect CSI at the BS and the user. However, the limited availability of time-frequency resources leads to issues like pilot contamination and inaccurate CSI information. In CFmMIMO, challenges such as power control and pilot assignment, due to their centralized

nature, become more pronounced [17–19]. Downlink power control algorithm is particularly challenging in CFmMIMO systems.

These challenges, while significant, provide ample research opportunities. This chapter sets the stage by outlining the system models for both technologies, highlighting the challenges faced in pilot resource allocation for massive MIMO and power control in CFmMIMO. The details of how our published works address these challenges will be explored in subsequent chapters.

While this thesis focuses on pilot overhead and its impact on the performance of massive MIMO systems, power control in massive MIMO is not addressed. This is because, although power control is important, it presents fewer challenges in massive MIMO compared to CFmMIMO, where the centralized coordination across distributed antennas makes power control more complex. As a result, the thesis prioritizes the investigation of pilot-related challenges in massive MIMO, while addressing power control in the context of CFmMIMO.

2.2 Massive MIMO Systems

2.2.1 System Description

We consider a multi-cell, multi-user massive MIMO system, hereafter simply referred to as massive MIMO. This system is characterized by L distinct cells, each having a BS equipped with M antennas serving K single-antenna users. The number of antennas M at each BS is significantly larger than the number of users K , facilitating spatial multiplexing, where all the users use the same time-frequency resources for communication.

The system employs a time-varying and frequency-selective block-fading channel model. The model assumes invariance over a coherence bandwidth, B_c , and a coherence time, T_c , creating a coherence block with $\tau_c = B_c T_c$ symbols. Here, the coherence bandwidth and time represent, respectively, the range of frequencies and the duration for which the channel's response is considered flat.

The BS performs essential operations like receive combining in the UL and transmit precoding in the DL. The receive combining module coherently combines signals from a given user to different receiver antennas thereby enhances the Signal-to-Noise Ratio (SNR) and minimizes inter-user interference. On the other hand, transmit precoding pre-adjusts signals from different antennas to superimpose constructively at the target user's location, boosting signal quality and mitigating multi-user interference in the downlink. The computational burden of these complex operations is delegated to the BS due to its superior computational capabilities and resources, allowing user devices to be less complex and more power-efficient.

2.2.2 TDD for Massive MIMO

Accurate CSI is necessary for the implementation of receive combining and transmit precoding. As the number of antennas M grows, the number of pilot symbols needed for DL channel estimation increases, leading to increased overhead and complexity. This situation is problematic, given that these pilot symbols consume valuable resources which could be otherwise used for data transmission. The available transmission resources for the pilot are constrained by the coherence time and bandwidth. Furthermore, DL CSI acquisition process induces a heavier computational load on user devices and demands more resources for CSI feedback in the uplink.

Time division duplexing (TDD) is often chosen to navigate this complex scenario due to its inherent property of channel reciprocity in the UL and DL. In a TDD system, the same frequency band is used for both UL and DL transmissions but at different times. Consequently, the DL channel can be estimated from the UL channel, alleviating the overhead associated with DL channel estimation and the CSI feedback in UL. A crucial advantage of this approach is that both the uplink pilot sequence length and the CSI feedback overhead scale only with the number of users within a cell, and not with the number of BS antennas. This guarantees that the channel estimation overhead remains manageable, even as the number of BS antennas increases, ensuring system scalability.

To leverage the reciprocity property, the TDD system divides the time within a coherence block into distinct periods for UL and DL transmissions. At the start of each coherence block, users transmit their pilot signals in the UL. Following this, the users transmit their data in the UL. The BSs utilize the UL pilots to determine the UL channel, which is then used to decode the UL data via the receive combining process. During the subsequent DL transmission phase, the BSs perform precoded DL data transmission using the estimated UL channel, thanks to channel reciprocity.

2.2.3 System Model

The UL channel between the user (l, k) (the k^{th} user in the l^{th} cell) and the BS in the j^{th} cell (BS j) is represented as $\mathbf{h}_{jlk} \in \mathbb{C}^M$. This channel follows a complex normal distribution $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{jlk})$, with \mathbf{R}_{jlk} being the spatial covariance matrix that encapsulates the statistical properties of the channel. Here, $\mathbf{0}$ stands for that the mean of all the elements of the channel being zero.

As elaborated in Subsection 2.2.2, accurate channel estimation is vital for implementing receive combining and transmit precoding. In TDD mode, the users transmit UL pilots to perform channel estimation. Consequently, massive MIMO dedicates a set of P ($\geq K$) symbols to UL pilots for channel estimation in each coherence block. Suppose the pilot sequence $\mathbf{p}_k \in \mathbb{C}^P$ is used by the k^{th} user of all L cells for channel estimation. For a different user q , we assume that the pilot sequence \mathbf{p}_q is mutually orthogonal to \mathbf{p}_k , and unit power is allocated

to all the users, i.e., $\mathbf{p}_k^H \mathbf{p}_q = P\delta_{kq}$. Further, we assume that the same P pilots are reused in each cell. UL transmit power is denoted by μ .

Assuming perfect synchronization in pilot transmissions across all cells, the signal received at BS j during pilot transmissions in the n^{th} coherence block, denoted as $\mathbf{Y}_j^{(p)}[n]$, can be expressed as:

$$\mathbf{Y}_j^{(p)}[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\mu} \mathbf{h}_{jlk} \mathbf{p}_k^T + \mathbf{N}_j^{(p)}[n]. \quad (2.1)$$

In this equation, $\mathbf{N}_j^{(p)}[n] \in \mathbb{C}^{M \times P}$ represents the noise during the pilot transmission. The noise elements are mutually independent and follow the distribution $\mathcal{CN}(0, 1)$. The superscript (p) indicates that the signals are related to the pilot transmission phase.

Subsequently, the UL data that is transmitted by user (l, k) in the same coherence block n is denoted as $\mathbf{x}_{lk}[n] \in \mathbb{C}^{C_u}$. Here, C_u is the number of symbols allocated for UL data transmission in each coherence block. This data is assumed to be a complex Gaussian vector with zero mean and identity covariance matrix, expressed as $\mathbf{x}_{lk}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

The signal received at BS j , denoted as $\mathbf{Y}_j[n] \in \mathbb{C}^{M \times C_u}$, consists of superposition of all received signals from all the users in the n^{th} coherence block, perturbed by AWGN noise $\mathbf{N}_j[n] \in \mathbb{C}^{M \times C_u}$. The noise elements are mutually independent and follow the unit normal distribution. The signal $\mathbf{Y}_j[n]$ can be represented as:

$$\mathbf{Y}_j[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\mu} \mathbf{h}_{jlk} \mathbf{x}_{lk}^T + \mathbf{N}_j[n]. \quad (2.2)$$

In contrast with (2.1), the superscript (p) is skipped in the above equation, indicating that these signals correspond to the data transmission phase.

During the DL communication, the base BS l conveys payload data, denoted as $\mathbf{d}_{lk}[n] \in \mathbb{C}^{C_d}$, to its associated user (l, k) , where C_d is the number of symbols allocated for DL data transmission in each coherence block. This data, which follows a complex normal distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I})$, is transmitted via a precoding vector $\mathbf{b}_{lk} \in \mathbb{C}^M$. The precoding vector is normalized to ensure that the average transmitted power, represented by λ , remains constant, that is, $\mathbb{E}\{\|\mathbf{b}_{lk}\|^2\} = 1$.

The signal received $\mathbf{z}_{ju}[n] \in \mathbb{C}^{C_d}$ at user (j, u) is distorted by AWGN noise $\mathbf{e} \in \mathbb{C}^{C_d}$, assumed to be complex Gaussian noise with zero mean and identity covariance matrix, i.e., $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. This can be formulated as below.

$$\mathbf{z}_{ju}[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\lambda} (\mathbf{h}_{lju}^H \mathbf{b}_{lk}) \mathbf{d}_{lk}[n] + \mathbf{e}_{ju}[n].$$

2.2.4 Pilot Contamination

The time-varying and frequency-selective nature of wireless channels imposes limitations on the available pilot resources, which are bounded by B_c and T_c .

Allocating more pilot resources increases overhead and reduces SE, while using fewer pilot resources can result in pilot contamination. Pilot contamination leads to issues such as inaccurate channel estimation and improper design of transmit precoders and receive combiners, resulting in degraded system performance [1, 20, 21]. This presents a trade-off between increasing pilot resources, which reduces SE due to overhead, and using fewer pilots, which increases pilot contamination. In Chapter 3, this trade-off is further explored with a focus on the overhead related to covariance estimation, supported by simulation results.

Pilot contamination, a fundamental challenge in massive MIMO, occurs when the pilot sequences assigned to different users are non-orthogonal, leading to inter-user interference. Within the same cell, pilot contamination is prevented through orthogonal pilot allocation, leveraging the assumption that $P \geq K$. However, due to the limited availability of pilot resources, pilot reuse across users in different cells becomes unavoidable. In the following subsections, we review the widely adopted LMMSE channel estimation technique, a robust method in the presence of inter-cell interference caused by the pilot reuse.

2.2.5 Channel Estimation

The LMMSE channel estimation technique aims to estimate the channels by utilizing the LS channel estimate and considering the covariance matrices of the target user's and interfering users' channels. Let the LS channel estimate of user (j, u) at BS j in the n^{th} coherence block be $\hat{\mathbf{h}}_{jju}^{LS}[n]$. It is obtained by minimizing the Euclidean distance between the received signal during pilot transmission, $\mathbf{Y}_j^{(p)}[n]$, and the estimated signal $\sqrt{\mu}\mathbf{g}\mathbf{p}_u^T$, where \mathbf{g} is the parameter vector for optimization. Solving this simple linear optimization problem, the LS channel estimate can be derived to obtain the following expression.

$$\hat{\mathbf{h}}_{jju}^{LS}[n] = \frac{1}{P\sqrt{\mu}} \mathbf{Y}_j^{(p)}[n] \mathbf{p}_u^* = \mathbf{h}_{jju} + \sum_{l \neq j} \mathbf{h}_{lju} + \frac{1}{P\sqrt{\mu}} \mathbf{N}_j^{(p)}[n] \mathbf{p}_u^*. \quad (2.3)$$

Then, as shown in [22], the corresponding LMMSE estimate $\hat{\mathbf{h}}_{jju}^{LMMSE}[n]$ is given by:

$$\hat{\mathbf{h}}_{jju}^{LMMSE}[n] = \mathbf{R}_{jju} \mathbf{Q}_{ju}^{-1} \hat{\mathbf{h}}_{jju}^{LS}[n]. \quad (2.4)$$

Here, the matrix \mathbf{Q}_{ju} is given by $\mathbb{E}\{\hat{\mathbf{h}}_{jju}^{LS}[n](\hat{\mathbf{h}}_{jju}^{LS}[n])^H\}$ and can be expressed as $\mathbf{Q}_{ju} = \sum_{l=1}^L \mathbf{R}_{lju} + \frac{1}{P\mu} \mathbf{I}$.

2.2.6 Spectral Efficiency Analysis for Uplink and Downlink

This subsection focuses on the exploration of SE within the massive MIMO architecture. The study utilizes a linear receive combiner and a linear transmit precoder for UL and DL communications, respectively.

In the context of UL, the linear receive combiner for user (j, u) at BS j is represented as $\mathbf{v}_{ju} \in \mathbb{C}^M$. By assuming LMMSE channel estimation, the tightest lower bound for UL channel capacity of user (j, u) within a massive MIMO system is given by [7]:

$$SE_{ju}^{(ul)} = \left(1 - \frac{P}{C_u}\right) \mathbb{E} \log_2 \left(1 + \gamma_{ju}^{(ul)}\right), \quad [\text{bits/s/Hz}]. \quad (2.5)$$

The pre-log factor in this equation accounts for pilot overhead, and the instantaneous Signal-to-Interference-plus-Noise Ratio (SINR), $\gamma_{ju}^{(ul)}$, is expressed as:

$$\gamma_{ju}^{(ul)} = \frac{|\mathbf{v}_{ju}^H \hat{\mathbf{h}}_{jju}|^2}{\mathbf{v}_{ju}^H \left(\sum_{l=1}^L \sum_{\substack{k=1 \\ (l,k) \neq (j,u)}}^K \hat{\mathbf{h}}_{jlk} \hat{\mathbf{h}}_{jlk}^H + \mathbf{Z}_j^{ul} \right) \mathbf{v}_{ju}}, \quad (2.6)$$

where, $\mathbf{Z}_j^{ul} = \sum_{l=1}^L \sum_{k=1}^K (\mathbf{R}_{jlk} - \mathbf{R}_{jlk} \mathbf{Q}_{lk}^{-1} \mathbf{R}_{jlk}) + \frac{1}{\lambda} \mathbf{I}$. Note that we have excluded the LMMSE superscript and the coherence block index n of the channel estimate for simplicity.

Moving on to DL data transmission, as described in Subsection 2.2.3, the linear precoding vector for user (l, k) at BS l is represented as \mathbf{b}_{lk} . Analogously to the UL SE, a lower bound for DL channel capacity of user (j, u) within a massive MIMO system is given by [7]:

$$SE_{ju}^{(dl)} = \mathbb{E} \log_2 \left(1 + \gamma_{ju}^{(dl)}\right), \quad [\text{bits/s/Hz}]. \quad (2.7)$$

The instantaneous SINR for DL is then given by:

$$\gamma_{ju}^{(dl)} = \frac{|\mathbf{b}_{ju}^H \hat{\mathbf{h}}_{jju}|^2}{\left(\sum_{l=1}^L \sum_{\substack{k=1 \\ (l,k) \neq (j,u)}}^K \mathbf{b}_{lk}^H \hat{\mathbf{h}}_{lju} \hat{\mathbf{h}}_{lju}^H \mathbf{b}_{lk} + z_j^{dl} \right)}, \quad (2.8)$$

where $z_j^{dl} = \sum_{l=1}^L \sum_{k=1}^K \mathbf{b}_{lk}^H (\mathbf{R}_{lju} - \mathbf{R}_{lju} \mathbf{Q}_{ju}^{-1} \mathbf{R}_{lju}) \mathbf{b}_{lk} + \frac{1}{\mu}$.

In TDD systems, the DL communication relies on channel hardening property and long-term statistical knowledge of the DL channel. This property is supported by the fact that the instantaneous channel gains closely approximate their statistical averages. The achievable SE expressions presented in this subsection are formulated by taking this property into consideration and invoking the worst-case Gaussian distribution technique [20, 23].

2.2.7 Imperfect CSI: Challenges in Pilot Resource Allocation

Assuming asymptotic linear independence of interfering users' covariance matrices, [24, 25] demonstrated that the pilot contamination is not a fundamental

asymptotic limitation in massive MIMO. In simpler terms, the sum of UL SE (given in (2.5)) of all the users in a cell continues to grow boundlessly with M .

The assumption of LMMSE channel estimation is essential for the validity of the UL SE bound given in (2.5). In addition to LMMSE channel estimation assumption, the assumption of LMMSE receive combining is also vital for the asymptotic boundlessness result in [24, 25]. However, it should be noted that the LMMSE channel estimation requires perfect knowledge of covariance matrices, which is often absent in practical massive MIMO systems.

The challenge lies in estimating individual covariance matrices in practical settings where raw channel estimates, from which the covariance estimates are computed, are themselves contaminated, introducing additional pilot overhead to achieve the necessary accuracy. This creates a trade-off between better covariance estimation and the impact on SE due to the increased pilot overhead. Addressing this trade-off is essential for managing system performance.

The covariance estimation problem has been extensively investigated in studies such as [22, 26–30]. These studies assume channel covariance matrices to be constant across multiple coherence blocks and use observations from a few of these blocks to estimate the covariance matrices.

The work [26] performs user clustering of multipath components using the power-delay function to estimate the spatial covariance matrices, while [29] presents a technique that allocates varying pilots across different coherence blocks to attain maximum-likelihood estimates of the covariance matrices. Moreover, [22] suggests two methodologies to prevent contamination in covariance matrices by dedicating additional orthogonal pilots to each user. Similarly, [30] introduces additional pilots for covariance estimation, but with a unique pilot structure, and it also develops corresponding covariance estimation method. The additional pilots for this method are not exclusively dedicated to each user, like in [22]. As a result, the number of extra pilots in [30] does not increase with the total number of users in the entire system (LK).

The method proposed in [30] offers higher throughput and lower mean squared error (MSE) of the (resulting) channel estimates compared to the method in [22]. Despite the need for additional pilots, this method refrains from assuming additional structures on the true covariance matrices of the users, unlike [26–28]. Moreover, this method does not require backhaul communication between neighboring cells, as in [29]. Therefore, this thesis will explore the performance of massive MIMO with the covariance estimation method outlined in [30].

In Chapter 3, we present a detailed description of the structure of the additional pilots and the corresponding covariance estimation method proposed in [30]. We also present the contributions of publications Publication I and Publication II in providing SE expressions (averaged over estimated covariance matrices) for massive MIMO with imperfect CSI, particularly, imperfect covariance information. It is important to note that the SE expressions given in (2.5) and (2.7) are no longer valid due to the violation of the perfect CSI knowledge assumption in LMMSE channel estimation.

The chapter then describes how these publications utilize the newly derived SE expressions to determine the required amount of additional pilot overhead for UL and DL communication. It also presents the SE expressions for a computationally efficient variant of the covariance estimation, where only the diagonal elements of the covariance are estimated, as derived in these publications. Based on numerical simulations, the chapter presents the conclusions drawn from these publications.

The newly derived SE expressions serve as crucial tools for quantifying the trade-off between better covariance estimation and the impact of additional pilot overhead on SE. By analyzing these expressions, it becomes possible to determine the optimal pilot resource allocation to balance estimation accuracy and reduced pilot overhead. Chapter 3 provides a detailed analysis of this trade-off, addressing the core problem statement introduced in this section.

2.3 Cell-Free Massive MIMO Systems

2.3.1 System Description

The CFmMIMO system under consideration consists of M access points (APs) or BSs uniformly distributed over a wide coverage area, serving K single-antenna users. Each AP is equipped with N antennas. Unlike traditional massive MIMO systems, CFmMIMO eliminates the concept of distinct cells or cell-boundaries. Instead, all APs are interconnected to a Central Processing Unit (CPU) via a fronthaul network, which orchestrates them to simultaneously serve all users using the same time-frequency resources.

We assume a block fading channel similar to our previous massive MIMO setup, with τ symbols within a coherence block. Furthermore, the CFmMIMO system operates in TDD mode, as explained in Subsection 2.2.2. This mode is primarily chosen for its ability to estimate UL and DL channels based solely on UL pilots, leveraging channel reciprocity. In this TDD configuration, each coherence block is partitioned into three transmission time intervals: 1. UL training, 2. UL data transmission, and 3. DL data transmission.

When it comes to communication techniques, the CFmMIMO system uses conjugate beamforming for both UL and DL communication. This technique is favored for its computational simplicity, which allows for UL channel estimation and UL and DL beamforming to be conducted at the APs in a distributed manner, thus relieving the CPU of a considerable computational load.

In this setup, the CPU performs slow rate processing that does not require instantaneous CSI. It only receives the payload data from the APs. Most of the algorithms that require instantaneous CSI, such as UL channel estimation and UL and DL beamforming, are executed at the APs. This distributed processing capability of the CFmMIMO system is effectively utilized. Consequently, the

CPU handles global level processing tasks such as power control and pilot assignment, which can significantly improve the performance of the CFmMIMO system without the need of instantaneous CSI.

2.3.2 System Model

The channel linking the k^{th} user to the m^{th} AP is characterized by $\mathbf{g}_{mk} = \beta_{mk}^{1/2} \mathbf{h}_{mk}$. In this expression, β_{mk} denotes the large-scale fading coefficient (LSFC), and $\mathbf{h}_{mk} \in \mathbb{C}^N$ represents the small-scale fading coefficients for each antenna of the AP, following a distribution of $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. We assume that this channel remains constant within a coherence block. In the following sections, we provide detailed descriptions of the uplink training (including channel estimation) and downlink data transmission (encompassing DL SE expressions) phases of operation, thereby laying foundation for the DL power control problem statement.

Uplink Training and Channel Estimation

In each coherence block, a segment of symbols, denoted by τ_p ($\ll \tau$), is reserved for pilot transmission. The k^{th} user transmits the pilot sequence, $\sqrt{\tau_p} \boldsymbol{\psi}_k \in \mathbb{C}^{\tau_p}$, with the constraint $\|\boldsymbol{\psi}_k\|_2 = 1$. The noise matrix, represented by $\mathbf{Z}_{p,m} \in \mathbb{C}^{N \times \tau_p}$, consists of elements that are independently and identically distributed as per $\mathcal{CN}(0, 1)$. The uplink received signal at the m^{th} AP is hence expressed as

$$\mathbf{R}_m = \sqrt{\zeta_p \tau_p} \sum_{i=1}^K \mathbf{g}_{mi} \boldsymbol{\psi}_i^T + \mathbf{Z}_{p,m},$$

where, ζ_p denotes the transmit SNR for each pilot symbol.

We assume the LSFCs are available at the APs and the CPU. Now, we can proceed to estimate the \mathbf{g}_{mk} using Minimum Mean Square Error (MMSE) estimator. The estimate is computed as

$$\hat{\mathbf{g}}_{mk} = \frac{\sqrt{\zeta_p \tau_p} \beta_{mk}}{1 + \zeta_p \tau_p \sum_{i=1}^K \beta_{mi} |\boldsymbol{\psi}_i^H \boldsymbol{\psi}_k|^2} \mathbf{R}_m \boldsymbol{\psi}_k^*. \quad (2.9)$$

Furthermore, the mean square value of the n^{th} element in $\hat{\mathbf{g}}_{mk}$ is denoted by v_{mk} , and is computed as

$$v_{mk} = \mathbb{E}(|\hat{\mathbf{g}}_{mk}[n]|^2) = \frac{\zeta_p \tau_p \beta_{mk}^2}{1 + \zeta_p \tau_p \sum_{i=1}^K \beta_{mi} |\boldsymbol{\psi}_i^H \boldsymbol{\psi}_k|^2}.$$

It is important to note that this value, v_{mk} , remains invariant to the antenna element index n .

Downlink Data Transmission

Here, we focus on matched filter beamforming. Let the downlink payload data symbol for the k^{th} user be c_k , such that $\mathbb{E}\{|c_k|\} = 1$. Let ζ_d be the maximum transmit SNR of each data symbol, normalized to the noise power. Let μ_{mi}

represent the power control coefficient of the signal directed to the i^{th} user from the m^{th} AP. Hence, the beamformed downlink signal at the m^{th} AP is constructed as follows:

$$\mathbf{x}_m = \sqrt{\zeta_d} \sum_{i=1}^K \frac{\mu_{mi}}{\sqrt{\nu_{mi}}} \hat{\mathbf{g}}_{mi}^* c_i.$$

It is noteworthy that the total transmit power at the m^{th} AP is given by $\mathbb{E}\{\|\mathbf{x}_m\|_2^2\} = \zeta_d N \sum_{i=1}^K \mu_{mi}^2$.

Let $z_{d,m}$ represent the noise signal, distributed as $\mathcal{CN}(0, 1)$. Therefore, the signal received by the k^{th} user is as follows:

$$r_k = \sum_{m=1}^M \mathbf{g}_{mk}^T \mathbf{x}_m + z_{d,m}. \quad (2.10)$$

We next consider the downlink spectral efficiency of a user within the CFm-MIMO system. Let us define $\boldsymbol{\mu}_m \triangleq [\mu_{m1}, \dots, \mu_{mK}]$ as the vector of power control coefficients associated with the m^{th} AP, and $\mathbf{M} \triangleq [\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_M^T]^T$ as the matrix of all power control coefficients. It is important to note that the power control coefficients are positive real numbers that satisfy the downlink power constraint $\mathbb{E}\{\|\mathbf{x}_m\|_2^2\} \leq \zeta_d$, or equivalently, $\|\boldsymbol{\mu}_m\|_2^2 = \sum_{i=1}^K \mu_{mi}^2 \leq 1/N$, for $1 \leq m \leq M$.

Let \mathbf{D}_k be a diagonal matrix with $\sqrt{\beta_{mk}}$ as the m^{th} diagonal element, and $\bar{\boldsymbol{\mu}}_i \triangleq [\mu_{1i}, \dots, \mu_{Mi}]^T$ be the vector of power control coefficients associated with the i^{th} user. Also, let

$$\mathbf{v}_{ik} \triangleq |\boldsymbol{\psi}_k^T \boldsymbol{\psi}_i^*| \left[\sqrt{\nu_{1i}} \frac{\beta_{1k}}{\beta_{1i}}, \dots, \sqrt{\nu_{Mi}} \frac{\beta_{Mk}}{\beta_{Mi}} \right]^T.$$

Using the well-known use-and-then-forget capacity bounding technique in the mMIMO and CFmMIMO literature [8, 23, 31, 32], we present the following SINR, along with a lower bound on downlink spectral efficiency (bits/s/Hz) of the k^{th} user:

$$\gamma_k(\mathbf{M}) = \frac{\zeta_d (\bar{\boldsymbol{\mu}}_k^T \mathbf{v}_{kk})^2}{\zeta_d \sum_{i=1, i \neq k}^K (\bar{\boldsymbol{\mu}}_i^T \mathbf{v}_{ik})^2 + \frac{\zeta_d}{N} \sum_{i=1}^K \|\mathbf{D}_k \bar{\boldsymbol{\mu}}_i\|_2^2 + \frac{1}{N^2}}, \quad (2.11)$$

$$SE_k^{(dl)}(\mathbf{M}) = \left(1 - \frac{\tau_p}{\tau}\right) \log_2(1 + \gamma_k(\mathbf{M})). \quad (2.12)$$

2.3.3 Power Allocation in Cell-Free Massive MIMO

The CFmMIMO system, as previously described, is designed to provide uniformly good service to all users within the system. This level of service is achievable due to the architectural simplicity of the considered system, in conjunction with distributed processing techniques such as channel estimation and matched filter beamforming. However, the global optimization problems associated with UL and DL power control, as well as pilot allocation, pose significant challenges.

This thesis primarily focuses on addressing the power control problem within the CFmMIMO system. Power control in a CFmMIMO system is complex due to the computational difficulty involved in solving the non-convex max-min fairness maximization problem [8, 31]. In this subsection, we explore the power control problem of the CFmMIMO system, treating it as a constrained optimization problem.

We formulate the max-min fairness maximization problem for the DL power control, subject to power constraints $\|\boldsymbol{\mu}_m\|_2^2 \leq \frac{1}{N}, 1 \leq m \leq M$. Let $S^{(dl)} \triangleq \{\mathbf{M} | \mathbf{M} \geq \mathbf{0}; \|\boldsymbol{\mu}_m\|_2^2 \leq \frac{1}{N}, 1 \leq m \leq M\}$. The constrained max-min fairness DL power control problem can then be formulated as

$$\begin{aligned} & \underset{\mathbf{M}}{\text{maximize}} && \min_{1 < k < K} SE_k^{(dl)}(\mathbf{M}) \\ & \text{subject to} && \mathbf{M} \in S^{(dl)}. \end{aligned} \tag{2.13}$$

SCA has been the standard technique for handling such types of quasi-convex problems, and these problems are typically solved using off-the-shelf second-order methods such as interior point methods [8, 32–34]. However, due to the computational complexity involved in the second-order methods used in SCA solvers, the first-order APG method was proposed in [31] to solve the optimization problem more efficiently. In Publication III, we propose an ANN-based unsupervised learning method to further reduce the computational complexity in performing the DL power control. Unsupervised learning is particularly advantageous as it eliminates the need for large labeled datasets that are difficult to generate in power control scenarios. In Publication IV, we propose an improvised variant for better DL power control performance. Chapter 4 provides details of the contributions in Publication III and Publication IV.

While Fully Connected Networks (FCNs) have been employed in previous works, they flatten the input data, leading to a loss of structural information crucial in LSFCs, mainly the inter-user relationships. Although some works have explored alternatives like Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs) to preserve such structures, these approaches face challenges with pilot contamination and scalability, limiting their effectiveness in large-scale CFmMIMO networks. In contrast, attention-based neural networks, specifically transformers, can effectively preserve this structure and handle systems of large sizes, as demonstrated in Publications III and IV. The proposed transformer-based models not only capture inter-user channel relationships but also incorporate pilot contamination data, offering scalability to large-scale networks.

2.4 Conclusion

This chapter sets the foundation for subsequent discussions by outlining the key principles and challenges of Massive MIMO and CFmMIMO systems in

wireless communication. Emphasis has been placed on the impact of pilot overhead on massive MIMO system performance and vitality and challenges involved in power control for CFmMIMO systems. The subsequent Chapter 3 will delve into the specifics of Publication I and Publication II, presenting a detailed exploration of the topics introduced here for massive MIMO. This will be followed by Chapter 4, focusing on Publication III and Publication IV for insights into the contributions of this thesis for the CFmMIMO technology.

3. Performance Analysis of Massive MIMO Systems with Pilot Resource Allocation

3.1 Introduction

As described in the previous chapter, realizing enhanced SE potential of massive MIMO systems requires the acquisition of accurate CSI at the BS. However, due to the limited size of the coherence block, obtaining accurate CSI is a non-trivial task, consequently affecting system performance. In particular, the issue of pilot contamination, arising from the reuse of pilot sequences by the users in different cells, poses significant challenges in estimating the channels, as well as the covariance matrices, of individual users. This contamination significantly impacts the SE of the massive MIMO system, hence necessitating careful consideration of pilot overhead in resource allocation strategies.

The motivation of this chapter is to analyze the trade-off between the SE and the additional pilot overhead required for covariance matrix estimation in massive MIMO systems. A thorough understanding of this trade-off enables us to develop efficient resource allocation strategies to optimize system performance in practical scenarios. The following sections delve into the details of this analysis, describing the interplay of various factors contributing to the system performance and the implications for pilot resource allocation in massive MIMO systems.

3.2 Channel Estimation

For channel estimation, we consider two types of UL pilots, namely, (i) ChEst: pilots for estimating the channel and (ii) CovEst: pilots for estimating the covariance matrix. Both ChEst pilots and CovEst pilots are assumed to be of length P symbols.

3.2.1 LMMSE-type Channel Estimation

In the Chapter 2, the LMMSE channel estimation was discussed under the assumption that the covariance information was readily known to the BS. However, this is not always the case in practical scenarios. In realistic environments, the covariance information must be estimated at the BS. It is thus essential to substitute the actual covariance matrices with their estimated counterparts, denoted as $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$. The resultant formulation of the ‘LMMSE-type’ channel estimate is given by:

$$\hat{\mathbf{h}}_{jju}[n] = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1} \hat{\mathbf{h}}_{jju}^{LS}[n]. \quad (3.1)$$

The total computational complexity required for evaluating (3.1) is given by $\mathcal{O}(M^3 + M^2N_R + M^2N_Q)$, where, N_R and N_Q represent the number of pilot sequences (samples) utilized to compute $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$, respectively.

3.2.2 Element-wise LMMSE-type Channel Estimation

As an alternative to LMMSE channel estimation, element-wise channel estimation is a popular method which holds an advantage in requiring fewer samples or pilots for covariance estimation that does not scale with M [35]. The expression for the element-wise LMMSE estimate of the channel is given by:

$$[\hat{\mathbf{h}}_{jju}^{\text{el-LMMSE}}[n]]_p = \frac{[\mathbf{S}_{jju}]_{pp}}{[\mathbf{P}_{ju}]_{pp}} [\hat{\mathbf{h}}_{jju}^{LS}[n]]_p, \quad p \in \{1, \dots, M\} \quad (3.2)$$

where \mathbf{S}_{jju} and \mathbf{P}_{ju} are defined as the diagonalized versions of matrices \mathbf{R}_{jju} and \mathbf{Q}_{ju} , respectively. The corresponding element-wise LMMSE-type estimate with estimated covariance matrices, denoted as $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$, is represented as bellow

$$[\hat{\mathbf{h}}_{jju}^{\text{el}}[n]]_p = \frac{[\hat{\mathbf{S}}_{jju}]_{pp}}{[\hat{\mathbf{P}}_{ju}]_{pp}} [\hat{\mathbf{h}}_{jju}^{LS}[n]]_p, \quad p \in \{1, \dots, M\} \quad (3.3)$$

Each diagonal element of $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$ is derived from a sample variance estimator corresponding to the element of the channel vector. Considering N_R and N_Q as the number of samples for estimating $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$, respectively, the computational complexity for evaluating $\hat{\mathbf{h}}_{jju}^{\text{el}}$ is $\mathcal{O}(MN_R + MN_Q)$. Notably, while the element-wise LMMSE-type channel estimation significantly mitigates computational complexity, this comes with a trade-off: potential performance degradation arising from the omission of non-diagonal elements of $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$.

3.2.3 Covariance Matrix Estimation

In both LMMSE-type and element-wise LMMSE-type channel estimation methodologies, Publications I and II adopt the covariance estimator detailed

in [30]. This subsection provides an overview of the covariance estimation method from [30]. The rationale behind selecting this particular covariance estimation method was elaborated upon in the preceding chapter.

It is generally assumed that the spatial covariance matrices for all channels remain constant over a time-interval and bandwidth significantly longer than that of a single coherence block [22, 26, 29, 30, 36]. Therefore, we assume that the covariance matrices remain unchanged across a time-interval denoted by T_s and a system bandwidth B_s . This suggests that the matrices can be considered constant over $\tau_s = B_s T_s / \tau_c$ coherence blocks.

While each of the τ_s coherence blocks contains ChEst pilots dedicated for channel estimation, only N_R coherence blocks encompass additional CovEst pilots. The primary objective of these CovEst pilots is to enable the estimation of the pair \mathbf{R}_{jju} and \mathbf{Q}_{ju} , or alternatively $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$, for each series of τ_s contiguous coherence blocks.

To estimate \mathbf{Q}_{ju} , the sample covariance estimates are derived from the LS channel estimates spanning N_Q coherence blocks. Similarly, element-wise sample variance estimates are employed to obtain $\hat{\mathbf{P}}_{ju}$.

To estimate \mathbf{R}_{jju} and \mathbf{S}_{jju} , the CovEst pilot sequence, comprising P symbols transmitted by each user, is employed. Specifically, the CovEst pilot transmitted by user (l, k) within the n^{th} coherence block is represented as $\phi_{lk}[n] = e^{j\theta_{ln}} \mathbf{p}_k$, a phase-shifted variant of the ChEst pilots. The phase-shifts, represented as $\{\theta_{ln}\}_{n=1}^{N_R}$, conform to the uniform distribution $\mathcal{U}[0, 2\pi)$. A critical observation is that the CovEst pilots are transmitted in only N_R out of the τ_s coherence blocks, and they vary across different coherence blocks.

The LS channel estimates derived from the pilots \mathbf{p}_u and ϕ_{ju} are denoted as $\hat{\mathbf{h}}_{jju}^{(1)}[n]$ and $\hat{\mathbf{h}}_{jju}^{(2)}[n]$, respectively. They are defined as:

$$\hat{\mathbf{h}}_{jju}^{(1)}[n] = \hat{\mathbf{h}}_{jju}^{LS}[n] = \mathbf{h}_{jju} + \sum_{l \neq j} \mathbf{h}_{jlu} + \frac{1}{P\sqrt{\mu}} \mathbf{N}_j^{(p)}[n] \mathbf{p}_u^* \quad (3.4)$$

$$\hat{\mathbf{h}}_{jju}^{(2)}[n] = \mathbf{h}_{jju} + \sum_{l \neq j} \mathbf{h}_{jlu} e^{j(\theta_{ln} - \theta_{jn})} + \frac{1}{P\sqrt{\mu}} \mathbf{N}_j^{(r)}[n] \mathbf{p}_u^* e^{-j\theta_{jn}}. \quad (3.5)$$

Estimation of $\hat{\mathbf{R}}_{jju}$

The independence of the interference and noise terms (second and third terms) in (3.4) from those in (3.5) allows the (ensemble) cross-correlation between $\hat{\mathbf{h}}_{jju}^{(1)}[n]$ and $\hat{\mathbf{h}}_{jju}^{(2)}[n]$ to result in the covariance matrix \mathbf{R}_{jju} [30]. Consequently, the unbiased Hermitian-symmetric sample cross-covariance matrix can be expressed as an estimate for \mathbf{R}_{jju} :

$$\hat{\mathbf{R}}_{jju} = \frac{1}{2N_R} \sum_{n=1}^{N_R} \left(\hat{\mathbf{h}}_{jju}^{(1)}[n] \left(\hat{\mathbf{h}}_{jju}^{(2)}[n] \right)^H + \hat{\mathbf{h}}_{jju}^{(2)}[n] \left(\hat{\mathbf{h}}_{jju}^{(1)}[n] \right)^H \right). \quad (3.6)$$

As N_R approaches infinity, the estimated covariance matrix converges in probability towards the actual covariance matrix, symbolized as $\hat{\mathbf{R}}_{jju} \xrightarrow[N_R \rightarrow \infty]{P} \mathbf{R}_{jju}$.

Notably, for finite N_R , this unbiased estimator does not ensure positive diagonal elements. Thus, we introduce a regularized estimate for the covariance matrix:

$$\hat{\mathbf{R}}_{jju} = \alpha_R \check{\mathbf{R}}_{jju} + (1 - \alpha_R) \mathbf{R}_b, \quad (3.7)$$

where, \mathbf{R}_b represents an arbitrary symmetric positive definite bias-matrix, with α_R as a design parameter.

Estimation of \mathbf{S}_{jju}

For the purpose of element-wise LMMSE-type estimation, our focus is centered on estimating the diagonal matrix \mathbf{S}_{jju} . To achieve this, we employ an unbiased Hermitian-symmetric variance estimate, denoted as $\check{\mathbf{S}}_{jju}$:

$$[\check{\mathbf{S}}_{jju}]_{pp} = \frac{1}{2N_R} \left(\sum_{n=1}^{N_R} [\hat{\mathbf{h}}_{jju}^{(1)}[n]]_p [\hat{\mathbf{h}}_{jju}^{(2)}[n]]_p^* + [\hat{\mathbf{h}}_{jju}^{(2)}[n]]_p [\hat{\mathbf{h}}_{jju}^{(1)}[n]]_p^* \right) \quad \forall p \in \{1, \dots, M\}. \quad (3.8)$$

Further refining our approach, we introduce a regularized estimate for \mathbf{S}_{jju} , expressed as:

$$\hat{\mathbf{S}}_{jju} = \alpha_R \check{\mathbf{S}}_{jju} + (1 - \alpha_R) \text{diag}(\mathbf{R}_b). \quad (3.9)$$

3.3 Spectral Efficiency Metrics

The SE formulations described in Subsection 2.2.6 represent the tightest UL and DL SE bounds for massive MIMO systems. However, when the receivers adopt LMMSE-type or element-wise LMMSE-type channel estimators, these bounds no longer hold. For these cases where estimated channels use imperfect covariance information, we refer to relatively looser "use-and-then-forget" SE bounds, as given in [22, 29, 30]. The name "use-and-then-forget" derives from the underlying assumption that receivers utilize the channel estimates solely for receiver combining, bypassing their use for signal detection. These alternate bounds accommodate a range of channel estimation and combining methods.

3.3.1 Uplink Spectral Efficiency

Consider a receiver employing a maximum ratio combiner (MRC), represented by the combining vector $\mathbf{v}_{ju}[n] = \hat{\mathbf{h}}_{jju}[n] = \hat{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{jju}^{LS}[n]$, where $\hat{\mathbf{W}}_{ju}$ is given by:

$$\hat{\mathbf{W}}_{ju} = \begin{cases} \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}, & \text{for LMMSE-type channel estimate} \\ \hat{\mathbf{S}}_{jju} \hat{\mathbf{P}}_{ju}^{-1}, & \text{for element-wise LMMSE-type channel estimate.} \end{cases} \quad (3.10)$$

Upon combining, the received signal can be represented as:

$$\mathbf{v}_{ju}^H \mathbf{y}_j = \sqrt{\mu} \mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\} x_{ju} + \sqrt{\mu} (\mathbf{v}_{ju}^H \mathbf{h}_{jju} - \mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\}) x_{ju} \quad (3.11)$$

$$+ \sum_{k \neq u} \sqrt{\mu} \mathbf{v}_{ju}^H \mathbf{h}_{jkk} x_{jk} + \sum_{l \neq j} \sum_{k=1}^K \sqrt{\mu} \mathbf{v}_{ju}^H \mathbf{h}_{jlk} x_{lk} + \mathbf{v}_{ju}^H \mathbf{n}_j. \quad (3.12)$$

The terms in this expression correspond to various components of the received signal:

- The first term captures the desired signal component.
- The second term emerges from uncertainty in the array gain.
- The third term represents non-coherent intra-cell interference.
- The fourth term accounts for coherent interference due to pilot contamination.
- The final term represents additive noise.

Given that the desired signal component is uncorrelated with the other components, a lower bound on the SE of the UL channel from user (j, u) to BS j can be represented as [22]:

$$SE_{ju}^{(ul)} = \left(1 - \frac{P}{C_u} - \frac{N_R P}{C_u \tau_s}\right) \log_2 \left(1 + \gamma_{ju}^{(ul)}\right), \quad (3.13)$$

where the SINR, $\gamma_{ju}^{(ul)}$, is:

$$\gamma_{ju}^{(ul)} = \frac{|\mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\}|^2}{\sum_{l=1}^L \sum_{k=1}^K \mathbb{E}\{|\mathbf{v}_{ju}^H \mathbf{h}_{jlk}|^2\} - |\mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\}|^2 + \frac{1}{\mu} \mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{v}_{ju}\}}. \quad (3.14)$$

Within the pre-log factor, $\frac{P}{C_u}$ accounts for ChEst pilots, and $\frac{PN_R}{C_u \tau_s}$ accommodates for CovEst pilots.

For a particular realization of $\hat{\mathbf{W}}_{ju}$, the SINR expression can be further simplified as [22]:

$$\gamma_{ju}^{(ul)} = \frac{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})|^2}{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s) + \sum_{l=1}^L |\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2 - |\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})|^2}, \quad (3.15)$$

with

$$\mathbf{R}_s \triangleq \sum_{l=1}^L \sum_{k=1}^K \mathbf{R}_{jlk} + \frac{1}{\mu} \mathbf{I}. \quad (3.16)$$

3.3.2 Downlink Spectral Efficiency

This section presents the DL spectral efficiency for user (j, u) when employing a matched filter (MF) precoder. Specifically, the precoder can be represented as

$$\mathbf{b}_{ju} = \frac{\hat{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{ju}^{LS}}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{ju}^{LS}[n]\|^2\}}}.$$

For this scenario, the received signal at user (j, u) is formulated as:

$$\begin{aligned} z_{ju} = & \sqrt{\lambda} \mathbb{E}\{\mathbf{b}_{ju}^H \mathbf{h}_{ju}\} d_{ju} + \sqrt{\lambda} (\mathbf{b}_{ju}^H \mathbf{h}_{ju} - \mathbb{E}\{\mathbf{b}_{ju}^H \mathbf{h}_{ju}\}) d_{ju} \\ & + \sum_{k \neq u} \sqrt{\lambda} (\mathbf{b}_{ju}^H \mathbf{h}_{jk}) d_{jk} + \sum_{l \neq j} \sum_{k=1}^K \sqrt{\lambda} (\mathbf{b}_{ju}^H \mathbf{h}_{lk}) d_{lk} + e_{ju}. \end{aligned} \quad (3.17)$$

Each term in this expression represents distinct components:

- The first term signifies the desired signal component.
- The second term highlights the uncertainty in the DL transmit array gain.
- The third term denotes non-coherent intra-cell interference.
- The fourth term conveys coherent interference due to pilot contamination.
- Lastly, the fifth term denotes additive noise.

Assuming that the scalar in the denominator of the precoding vector, $\sqrt{\mathbb{E}\{\|\hat{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{ju}^{LS}[n]\|^2\}}$, is a constant known at the BS, we can derive the following SE expression. Similar to UL SE, the lower bound on DL channel SE for user (j, u) can be derived as:

$$\text{SE}_{ju}^{(dl)} = \log_2 \left(1 + \gamma_{ju}^{(dl)} \right) \quad [\text{bits/s/Hz}].$$

The SINR for the DL SE, in relation to a specific realization of $\hat{\mathbf{W}}_{ju}$, is:

$$\gamma_{ju}^{(dl)} = \frac{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{ju})|^2}{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s^{(dl)}) + \sum_{l=1}^L |\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{lu})|^2 - |\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{ju})|^2 + \frac{1}{\lambda}}, \quad (3.18)$$

$$\text{and } \mathbf{R}_s^{(dl)} \triangleq \sum_{l=1}^L \sum_{k=1}^K \mathbf{R}_{lk}.$$

3.4 Contributions

The performance of a massive MIMO system is closely linked to the quality of channel covariance estimates. The quality of these estimates largely depends on

the CovEst pilot overhead. Existing literature, such as [22, 29, 30], commonly focuses on the achievable SE for a single realization of covariance matrices, as demonstrated in equations (3.15) and (3.18). While this single-realization approach is crucial for practical receiver design, it does not sufficiently capture the effects of the CovEst pilot overhead. This observation underscores the necessity for average SE expressions, calculated over multiple covariance matrix realizations (ensemble average). Employing these expressions as performance metrics offers a comprehensive understanding of the impact of pilot overhead on system performance.

The main objective of the average SE-based performance metrics, detailed in Publications I and II, is to shed light on the effects of pilot overhead in massive MIMO systems. The focal point of this chapter is the foundational research presented in these two publications.

Publication I conducts a comprehensive exploration of the UL SE in massive MIMO systems. One of its pivotal contributions is the derivation of a closed-form expression for average UL SE under LMMSE-type channel estimation. This deepens our understanding of the effects of pilot overhead on system performance. Moreover, it sets the groundwork for subsequent studies. Building on the groundwork laid by Publication I, Publication II encompasses a more holistic analysis of both UL and DL SEs using both LMMSE-type and element-wise LMMSE-type channel estimation methods.

Collectively, these publications significantly enhance our understanding of how to optimize pilot resource allocation in massive MIMO systems. In the next section, we will delve deeper into these key contributions, emphasizing their relevance in the design considerations related to pilot overhead.

3.5 Average Spectral Efficiency and Impact of Pilot Overhead

It is essential to recognize that the UL and DL SE expressions specified in (3.15) and (3.18) treat $\hat{\mathbf{W}}_{ju}$ as a known matrix in the expectation operation. In this study, however, we regard it as a random matrix, dependent on the random realizations of $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$ (or $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$). Moreover, we assume mutual independence between $\hat{\mathbf{W}}_{ju}$ and $\hat{\mathbf{h}}_{jju}^{LS}[n]$. As such, $\mathbb{E}\{\cdot\} = \mathbb{E}_W\{\mathbb{E}_{h^{LS}}\{\cdot\}\}$, where \mathbb{E}_W represents the expectation over $\hat{\mathbf{W}}_{ju}$, and $\mathbb{E}_{h^{LS}}$ is the expectation over the LS estimate.

In Publications I and II, the UL and DL SINR expressions for this configuration are given by:

$$\gamma_{ju}^{(ul)} = \frac{|\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\}|^2}{\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s)\} + \sum_{l=1}^L \mathbb{E}_W\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2\} - |\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\}|^2} \quad (3.19)$$

and

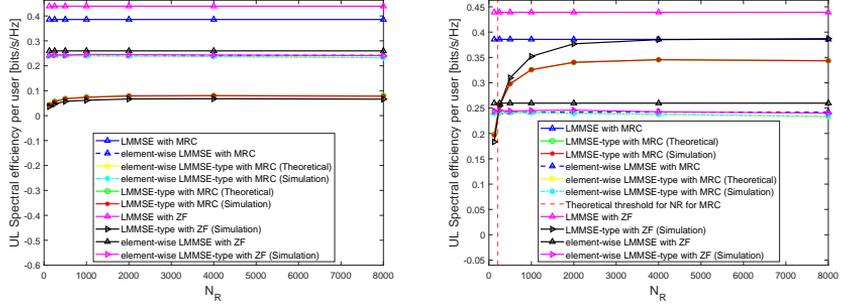
$$\gamma_{ju}^{(dl)} = \frac{|\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\}|^2}{\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s^{(dl)})\} + \sum_{l=1}^L \mathbb{E}_W\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2\} - |\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\}|^2 + \frac{1}{\lambda}}, \quad (3.20)$$

with subsequent derivations of the results of the expectation operation in each term embedded in these SINR expressions, respectively. When these expressions are evaluated for LMMSE-type and element-wise LMMSE-type channel estimation methods, they manifest as functions of N_R and N_Q . This formulation aids in analyzing the effect of CovEst pilots. For the sake of conciseness, we refrain from detailing the exact outcomes of these expressions, directing readers instead to Theorems 1 and 2 in Publication II.

Although the derived SE expressions in these theorems are for MRC combining and MF precoding, we also consider ZF combining and precoding in the simulations and analysis. Due to mathematical intractability of the expectation operation for ZF beamforming, we consider only numerical averages of the SE expressions. The work [37] demonstrated that the ZF beamforming along with the LMMSE channel estimation, when the covariance matrix is known, provides an effective combination. Therefore, here we present a comprehensive overview of the massive MIMO system performance, under imperfect covariance information, that is not only impacted by the pilot overhead, but also by the channel estimation choice (between LMMSE-type and element-wise LMMSE-type techniques) and beamforming choice (between MF and ZF beamforming).

Using (3.19) and (3.20), we have established a direct relation between the average SE value and the parameters N_R and N_Q . The expectation terms given in these theorems contain two components: (i) the component that corresponds to the true covariance information and (ii) a penalty component due to regularization of \mathbf{R}_{jju} estimate and due to covariance estimation error. If $\alpha_R = 1$, and as N_R and N_Q tend to infinity, the penalty components of the expectation terms vanish.

In Fig. 3.1 and 3.2, we consider UL and DL SE with both channel estimation methods and both beamforming techniques for different values of N_R and N_Q . These figures are plotted for $L = 7$ cells with $M = 100$ and $K = 10$ users per cell. The BSs are at a distance of 300 m apart from each other, and the users are uniformly spaced at a distance of 120 m from the BS in their cells. The angular spread of the channel cluster is assumed to be 20° , within which the received paths from a user are assumed to be uniformly distributed. We consider a 3GPP urban macro [38] scenario with a non-line-of-sight channel for simulating the path loss model. Furthermore, we also consider coherence block length as $C_u = 100$ symbols, pilot sequence length as $P = 10$, and the number of coherence blocks for which the channel covariance matrices are considered constant as $\tau_s = 25000$. Additionally, we choose $\alpha_R = 0.95$, and $\mathbf{R}_b = \mathbf{I}$. Sample averaging for all the expectation terms is computed using 2000 trials. Our simulations are


 (a) SE vs N_R with $N_Q = 125$.

 (b) SE vs N_R with $N_Q = 4000$.

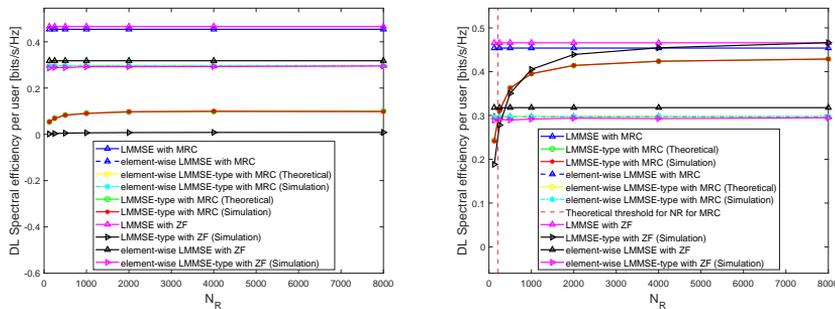
Figure 3.1. UL SE for different channel estimation techniques.

conducted for varying degrees of pilot overhead, quantified by different values of $N_R = \{125, 250, 500, 1000, 2000, 4000, 8000\}$ and $N_Q = \{125, 4000\}$.

Uplink Spectral Efficiency

In Fig. 3.1, we present a comparative analysis of UL SE curves for varying values of N_R . The methods evaluated include LMMSE channel estimation paired with MRC, element-wise LMMSE channel estimation (assuming perfect knowledge of the diagonal elements of the covariance matrix) paired with MRC, LMMSE-type channel estimation paired with both MRC and ZF combining, and lastly, the element-wise LMMSE-type channel estimation with both MRC and ZF. Additionally, the figure contrasts the theoretical SE curves for MRC combining with their simulated average SE counterparts. Specifically, Fig. 3.1(a) showcases the SE values when $N_Q = 125$, while Fig. 3.1(b) details the SE values for $N_Q = 4000$. Here is a summary of the analysis on UL SE curves:

- The simulated SEs match the theoretical values for both the channel estimation techniques tested, thereby validating the derivations.
- For MRC combining, while LMMSE is always better than element-wise LMMSE, LMMSE-type is not necessarily better than element-wise LMMSE-type.
- Using simulations for the given setup, we observe that element-wise LMMSE-type channel estimation outperforms LMMSE-type when N_Q is less than a threshold 263. Furthermore, for $N_Q = 4000$ (which is greater than 263), the element-wise LMMSE-type channel estimation outperforms the LMMSE-type for N_R below the derived theoretical threshold, \bar{N}_R .
- For both LMMSE-type and element-wise LMMSE-type channel estimations, when paired with MRC combining, the UL SE curve exhibits an initial increase


 (a) SE vs N_R with $N_Q = 125$.

 (b) SE vs N_R with $N_Q = 4000$.

Figure 3.2. DL SE for different channel estimation techniques.

with N_R due to improved covariance estimates. However, it subsequently declines due to the significant impact of CovEst pilot overhead.

- Compared to LMMSE-type, the SE for element-wise LMMSE-type reaches the SE for known covariance case faster.
- ZF combining performs well only for the larger number of pilots ($N_Q = 4000$ and $N_R \geq 500$) and needs additional computational complexity. However, the ZF combining does not significantly improve the performance for element-wise LMMSE channel estimation; it is marginally better than the performance corresponding to the MRC combining.
- For large N_R and N_Q values, ZF combining outperforms the MRC combining. This is enabled by the better covariance estimates.

Downlink Spectral Efficiency

Similar to the UL simulation, here, we consider the DL SE expressions corresponding to LMMSE channel estimation paired with MF precoding, element-wise LMMSE channel estimation (assuming perfect knowledge of the diagonal elements of the covariance matrix) paired with MF precoding, LMMSE-type channel estimation paired with both MF precoding and ZF precoding, and lastly, the element-wise LMMSE-type channel estimation with both MF precoding and ZF precoding. In Fig. 3.2, we plot the SE as a function of N_R for the two aforementioned channel estimation techniques. Fig. 3.2(a) depicts the SE values for $N_Q = 125$, and Fig. 3.2(b) shows SE values for $N_Q = 4000$. In DL SE plots we observe similar behavior to the UL SE case. Distinct from what we observe in UL SE, here, in DL SE we observe that the threshold for N_Q under which the element-wise LMMSE-type channel estimation always outperforms LMMSE-type channel estimation is 272.

4. Transformer Neural Network for Downlink Power Control in Cell-Free Massive MIMO Systems

4.1 Introduction

In Chapter 2, we outlined the CFmMIMO setup, its numerous advantages, and the critical role of power allocation in achieving optimal system performance. We also discussed the complexity challenges associated with downlink power control due to the non-convex nature of the problem and the large number of optimization parameters. These issues are compounded by the limitations of existing online iterative methods, which often fail to scale efficiently in large CFmMIMO networks.

To address these challenges, this chapter delves into advanced deep learning techniques, specifically transformers, for downlink power control. Leveraging unsupervised learning, these networks are trained to solve complex optimization problems without the need for extensive labeled data. Moreover, the proposed transformer neural networks effectively handle issues such as pilot contamination, which significantly impacts the performance of CFmMIMO systems. These methods offer a scalable and computationally efficient alternative to traditional online iterative approaches. The techniques discussed in this chapter are based on the methodologies presented in Publications III and IV, which contribute to the development and implementation of these neural networks. This chapter aims to provide a comprehensive description of these publications, highlighting their contributions to the field.

4.2 Unsupervised Learning Approach for Downlink Power Control

Recall that M is the number of APs or BSs in a CFmMIMO system simultaneously serving all the K users. Let us now define an $M \times K$ matrix of large-scale fading coefficients (LSFCs) as \mathbf{B} . The rows of \mathbf{B} correspond to the LSFCs of different APs, while the columns correspond to LSFCs of users. Note that, in Subsection 2.3.2, we have defined a similar $M \times K$ matrix of power control coeffi-

cients as \mathbf{M} . Let Φ be the $K \times K$ interference matrix, where the element located in the i^{th} row and k^{th} column is given by $\Phi_{ik} = |\boldsymbol{\psi}_k^T \boldsymbol{\psi}_i^*|^2$.

Let us consider an unsupervised learning setup utilizing an arbitrary Deep Neural Network (DNN) with trainable parameters \mathbf{W} , represented as $\mathbf{F}_{DNN}(\cdot, \cdot; \mathbf{W})$. The DNN takes the input pair (Φ, \mathbf{B}) and outputs $\mathbf{M} = \mathbf{F}_{DNN}(\Phi, \mathbf{B}; \mathbf{W})$. Furthermore, we define the soft-minimum utility function as:

$$u(\mathbf{M}; \Phi, \mathbf{B}) \triangleq -\frac{1}{\lambda} \ln \left(\frac{1}{K} \sum_{k=1}^K \exp(-\lambda SE_k(\mathbf{M}; \Phi, \mathbf{B})) \right). \quad (4.1)$$

The soft-minimum function provides a differentiable approximate of the hard minimum function, given by $\min_{1 \leq k \leq K} SE_k(\mathbf{M}; \Phi, \mathbf{B})$. Furthermore, in the above equation, λ is the smoothening parameter.

In the unsupervised learning framework, the goal is to maximize the expected smoothed-minimum utility function. The expectation operation attempts to train the DNN weights \mathbf{W} using several training examples, such that the average utility function, among the training samples, is maximized. During inference, this setup aims to replace expensive solvers for the max-min fairness problem, seeking the mapping $\mathbf{F}_D(\Phi, \mathbf{B}; \mathbf{W}_{\text{opt}})$ that achieves similar performance. The optimization problem that the training process solves is¹:

$$\underset{\mathbf{W}}{\text{maximize}} \quad \mathbb{E}[u(\mathbf{F}_{DNN}(\Phi, \mathbf{B}; \mathbf{W}); \Phi, \mathbf{B})], \quad (4.2)$$

where, the optimization parameter is \mathbf{W} instead of \mathbf{M} , unlike in (2.13).

Assuming the distribution of \mathbf{B} is unknown, we generate a large set of LSFC matrices $\{\mathbf{B}^{[p]} \in \mathbb{R}_+^{M \times K} \mid 1 \leq p \leq P\}$, representing different random user placements, based on a fixed AP placement and path loss model. Correspondingly, the pilot allocation algorithm produces $\{\Phi^{[p]} \in \mathbb{R}_+^{M \times K} \mid 1 \leq p \leq P\}$. Since the unsupervised learning setup described here is independent of the specifics of the data generation process, the details of the LFSCs and pilot information are deferred to Subsection 4.5.1. Using these sets, the optimal \mathbf{W} is found by maximizing the empirical average of the utility function:

$$\underset{\mathbf{W}}{\text{maximize}} \quad \frac{1}{P} \sum_{p=1}^P u(\mathbf{F}_{DNN}(\Phi^{[p]}, \mathbf{B}^{[p]}; \mathbf{W}); \Phi^{[p]}, \mathbf{B}^{[p]}). \quad (4.3)$$

As the number of samples P increases, the empirical average approximates the ensemble average, providing a practical alternative to the problem in (4.2). While the original ensemble problem cannot be solved directly, this empirical methodology offers a viable solution through extensive data-driven training.

¹Note that the optimization problem in (2.13) is distinct from the training problem given here.

4.3 Literature Review on Deep learning Based Solutions

Learning-based optimization solutions significantly reduce computational complexity by avoiding iterative solvers during real-time operations. These approaches efficiently map input LSFCs directly to the output power control coefficients, enhancing system performance and scalability [39–59].

For uplink power control, Fully Connected Network (FCN) based unsupervised learning solutions are discussed in [40–42], while supervised setups using Long Short-Term Memory (LSTM) and FCN are detailed in [43–45]. Various Reinforcement Learning (RL) based solutions are explored in [46–49].

For Downlink, FCN-based unsupervised learning methods are examined in [50–52], and RL-based solutions are detailed in [53]. Supervised Convolutional Neural Network (CNN) and Graph Neural Network (GNN) solutions are proposed in [54–56]. However, these methods often assume ideal conditions, overlooking pilot contamination and thereby encountering practical limitations in large-scale CFmMIMO systems. Although the studies in [57–59] consider pilot contamination scenarios while evaluation, they avoid handling pilot contamination scenarios while training or model design.

FCN methods tend to lose crucial associations among APs and users, while CNN methods preserve matrix \mathbf{B} 's structures but may not be suitable for LSFC matrices. GNN approaches utilize structural information effectively but do not adequately address pilot contamination complexities [56].

There remains a gap in DNN-based solutions for managing downlink power control in CFmMIMO systems faced with pilot contamination. A robust solution is needed to handle these complexities while maintaining the structural integrity of the LSFC matrix.

4.4 Contributions

This section presents the contributions made in two publications, referred to as Publication III and Publication IV. Both are built on modified transformer architectures. Publication III, a conference paper, and Publication IV, a journal article, both focus on solving the downlink power control problem in CFmMIMO systems using deep learning techniques.

Publication III introduces an attention neural network (ANN) designed to efficiently solve the power control problem by utilizing the structure in LSFCs and incorporating pilot allocation information as input. Building on this, Publication IV introduces the Pilot contamination-Aware Power Control (PAPC) transformer, an enhanced version of the ANN. It incorporates architectural improvements to demonstrate the transformer's ability to handle pilot contamination scenarios and offers scalability to larger-scale CFmMIMO networks. These contributions represent a significant advancement in solutions for downlink power control in CFmMIMO systems.

4.4.1 Why Transformer-Based Models for CFmMIMO?

The transformer architecture was chosen for downlink power control in CFmMIMO systems due to its ability to efficiently exploit the inherent structure in LSFCs. Unlike FCNs, which flatten input matrices and lose critical relational information, transformers preserve the two-dimensional structure of the LSFC matrix \mathbf{B} . By retaining the structure in the data, the transformer significantly reduces training time, lowers the amount of data required for training, and improves model accuracy.

Another key advantage of transformers is their masking mechanism, which, through customized modification, allows the incorporation of pilot allocation information. This is essential for handling both pilot-contaminated and contamination-free scenarios, as it enables the model to use the available prior pilot information to mitigate interference and optimize power control decisions.

4.4.2 Overview of ANN Model

The ANN model, proposed in Publication III, serves as the initial application of the transformer-based architecture for downlink power control in CFmMIMO systems. It leverages masked multi-head attention networks (MMHAN) to solve the challenging non-convex optimization problem of power control while maintaining the structural integrity of the LSFC matrix. This approach allows the ANN to capture inter-user channel dependencies effectively, a significant improvement over traditional FCNs.

The ANN model incorporates pilot allocation information, though at this stage of development, it was not explicitly evaluated in pilot contamination scenarios. Power constraints in the ANN are managed through a more intricate training process using an interior point method. The ANN demonstrated the potential of transformer solutions to improve computational complexity in the power control applications with reasonable performance close to state-of-the-art iterative solutions.

4.4.3 Transition to the PAPC Model (Publication IV)

The PAPC model, introduced in Publication IV, builds on the foundation of the ANN model and extends it with enhanced features and evaluations. The core transformer block, shown in Figure 4.1, remains the same as in the ANN model, consisting of a customized multi-head attention block (MMHA, the same as MMHAN referred to in Publication III). The overall architecture of the MMHA module is illustrated in Fig. 4.2. However, key improvements are made to the ANN using additional modules.

One of the major architectural enhancements in PAPC is the additional pre-processing stage, which increases the input dimensionality of the rows of the LSFC matrix. This allows the model to capture richer features. Additionally,

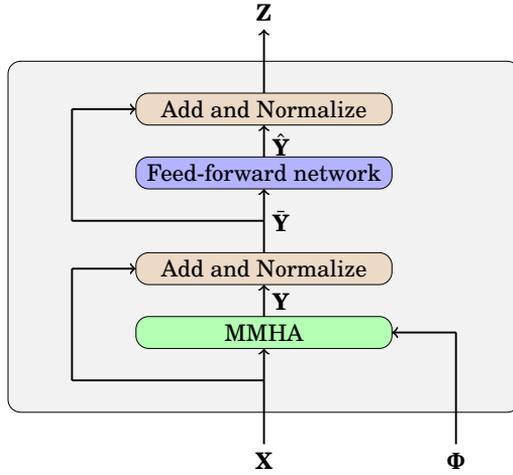
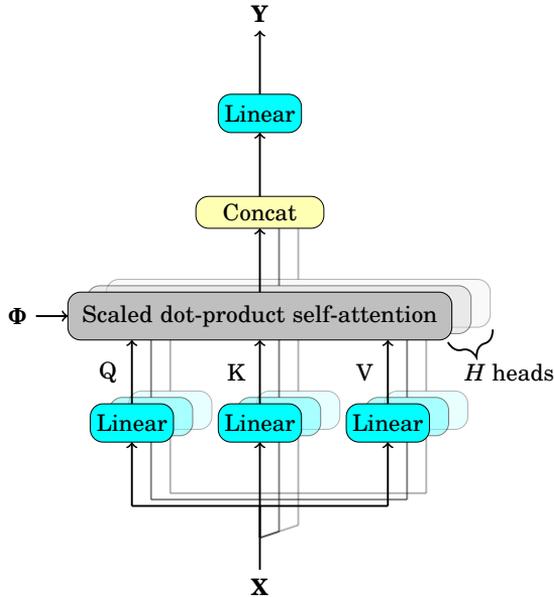


Figure 4.1. The PAPC transformer block processes the input using an MMHA and a feed-forward network, with residual connections and layer normalizations.

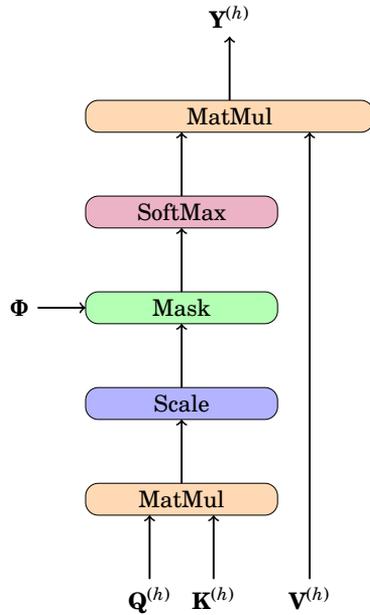
PAPC simplifies the handling of power constraints by replacing the complex interior-point method used in ANN with a more efficient projection operation in a new post-processing stage.

Furthermore, in Publication IV, the PAPC model was evaluated in more complex scenarios, including pilot contamination, which was not explored in the ANN model. This demonstrated PAPC's ability to effectively mitigate interference by leveraging pilot allocation information, resulting in improved SE and fairness. These evaluations highlight the robustness of PAPC in both pilot-contaminated and contamination-free environments, demonstrating its potential to efficiently utilize the available pilot information. PAPC's computational efficiency, scalability, and adaptability to varying number of users were also demonstrated in Publication IV.

The improved performance of the PAPC model can be attributed primarily to the enhanced feature extraction enabled by the additional preprocessing stage. This step allows the potential of the transformer-based architecture, initially proposed in ANN, to be fully realized. Since PAPC has demonstrated superior capability, the next section presents the results of the PAPC model. The performance of ANN is not included, as PAPC is both an enhancement of ANN and offers better overall performance.



(a) Overview of MMHA with multi-head attention and output combination.



(b) Scaled dot-product self-attention with custom masking.

Figure 4.2. MMHA architecture in the PACC transformer, processing the input through multiple attention heads combined with masking feature to model inter-user relationships and handle pilot contamination.

4.5 Results

4.5.1 Simulation Setup

To assess the effectiveness of the PAPC in downlink power control for CFm-MIMO systems, we analyze four scenarios labeled from Scenario 0 to Scenario 3. Scenario 0 represents a small-scale CFmMIMO network, consisting of $M = 10$ APs and $K = 4$ users within an area of 0.01 sq. kms. In contrast, Scenarios 1, 2, and 3 scale up the network to $M = 100$ APs, with $K = 20$, $K = 40$, and $K = 80$ users respectively, all distributed over 0.1 sq. kms. A wrap-around topology is employed to simulate a larger area and mitigate boundary effects. Across all scenarios, the performance evaluation is based on 2000 testing samples. Neural networks are trained using $P = 12,000,000$ samples unless otherwise stated. For a fair comparison of computational times across different algorithms, testing is carried out without GPU assistance.

In each of these scenarios, the AP density is maintained at 1000 APs per sq. km, with each AP featuring $N = 4$ antennas. The coherence block and pilot sequence lengths are defined as $\tau = 200$ symbols and $\tau_p = 20$ symbols, respectively. For our numerical simulations, we assume that the pilot allocation algorithm selects K sequences from a pool of τ_p orthogonal sequences. Consequently, in Scenarios 2 and 3, the number of users (K) surpasses the available orthogonal pilots (τ_p), necessitating pilot reuse. This scenario inevitably leads to pilot contamination, presenting a significant challenge in these large-scale configurations.

The distance between the m^{th} AP and the k^{th} user, denoted as d_{mk} in km, is used to determine the path loss, PL_{mk} , in dB for the channel connecting the m^{th} AP to the k^{th} user. This path loss is calculated using a three-slope model, given by $PL_{mk} = -L_0 - 15\log_{10}(d_1) - 20\log_{10}(d'_{mk})$. The parameter d'_{mk} is defined as:

$$d'_{mk} = \begin{cases} d_0 & \text{if } d_{mk} \leq d_0 \\ d_{mk} & \text{if } d_0 < d_{mk} \leq d_1 \\ d_1 & \text{if } d_{mk} > d_1. \end{cases} \quad (4.4)$$

The LSFC for the channel between the m^{th} AP and the k^{th} user is expressed as $\beta_{mk} = PL_{mk} + z_{mk}$ dB, where z_{mk} represents shadow fading, modeled as a normal distribution with a mean of 0 dB and a variance of σ_{sh}^2 dB. Following the setup in [8], the parameters are set to $L_0 = 140.72$ dB, $d_0 = 0.01$ km, $d_1 = 0.05$ km, and $\sigma_{sh} = 8$ dB.

Data generation process: Given a fixed placement of APs or BSs, the data generation process generates P random user placements in the given area of coverage. The process computes the LSFCs matrix \mathbf{B} sample using the path loss model described above for each user placement. Furthermore, the random pilot allocation algorithm described in this subsection produces corresponding pilot allocation matrix Φ .

| | |
|---|----------------------|
| AP Density | 2000 APs per sq. km. |
| Length of the coherence block | 200 symbols |
| Length of the pilot sequence | 20 symbols |
| L_0 | 140.72 dB |
| d_0 | 0.01 km |
| d_1 | 0.05 km |
| Standard deviation of shadow fading (σ_{sh}) | 8 dB |
| Noise power spectral density N_0 | -173.98 dBm/Hz |
| BandWidth | 20 MHz |
| Total Noise power at the receiver (P_n) | -91.97 dBm |
| Transmit SNR of uplink pilot (ζ_p) | $1/P_n$ |
| Transmit SNR of downlink data (ζ_d) | $0.2/P_n$ |
| Smoothing parameter (λ) | 3 |

Table 4.1. Simulation Setup Parameters

Given a noise figure of $N_f = 9$ dB, a noise power spectral density of $N_0 = -173.98$ dBm/Hz, and a channel bandwidth of $BW = 20$ MHz, the total noise power is computed as $P_n = BW10^{(N_0+N_f-30)/10}$ in W. Consequently, the transmit SNR for the uplink pilot and downlink data are set to $\zeta_p = 0.2/P_n$ and $\zeta_d = 1/P_n$, respectively. Table 4.1 outlines the common CFmMIMO simulation parameters for all scenarios.

4.5.2 Results Summary

- **Scenario 0 (Fig. 4.3):** As the number of training samples increases, both PAPC and FCN approach the performance of the APG algorithm. PAPC outperforms FCN with fewer training samples, highlighting its structural advantages. With sufficient samples, PAPC surpasses both FCN and EPA, matching the performance of APG.
- **Scenarios 1 to 3 (Fig. 4.4):** PAPC approaches APG performance across all scenarios, even in the presence of pilot contamination. FCN fails to perform adequately in Scenario 1 and deteriorates further with increased pilot contamination in Scenarios 2. PAPC's ability to handle pilot contamination and maintain performance is validated.
- **Computational Efficiency (Table 4.2):** PAPC significantly reduces computational time compared to APG, demonstrating its efficiency. PAPC's run-time is marginally higher than EPA but provides superior performance.
- **Varying K Feature (Fig. 4.5 and Fig. 4.6):** PAPC trained with varying K

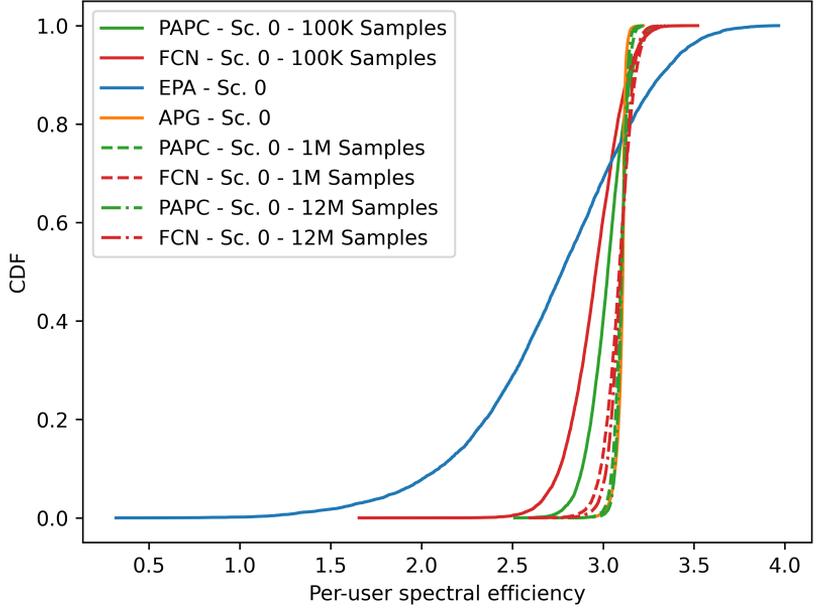


Figure 4.3. CDF comparison of PAPC, FCN, EPA, and APG in Scenario 0 for different training samples. To interpret the results, note that for the max-min fairness objective, a CDF curve that ascends sharply and is also positioned further to the right compared to other curves is considered advantageous. Thus, PAPC outperforms FCN and EPA, approaching APG's performance faster as the number of samples increases.

feature maintains performance across different user counts. Furthermore, the performance of PAPC with varying K matches that of PAPC trained with a fixed K , demonstrating adaptability without loss in performance.

Note that, unlike traditional optimization techniques, deep learning models do not offer mathematical guarantees of performance for specific configurations. Their correctness and efficiency can only be empirically validated through simulation results. This limitation makes it inherently challenging to provide a solid computational complexity analysis for the proposed methods. Moreover, since these models are typically tailored to specific CFmMIMO network configurations (e.g., number of users and access points), adapting to different configurations requires retraining or redesigning the network, further complicating a formal analysis. Nonetheless, as demonstrated in Table 4.2, the empirical results show that PAPC significantly reduces computational time compared to traditional methods.

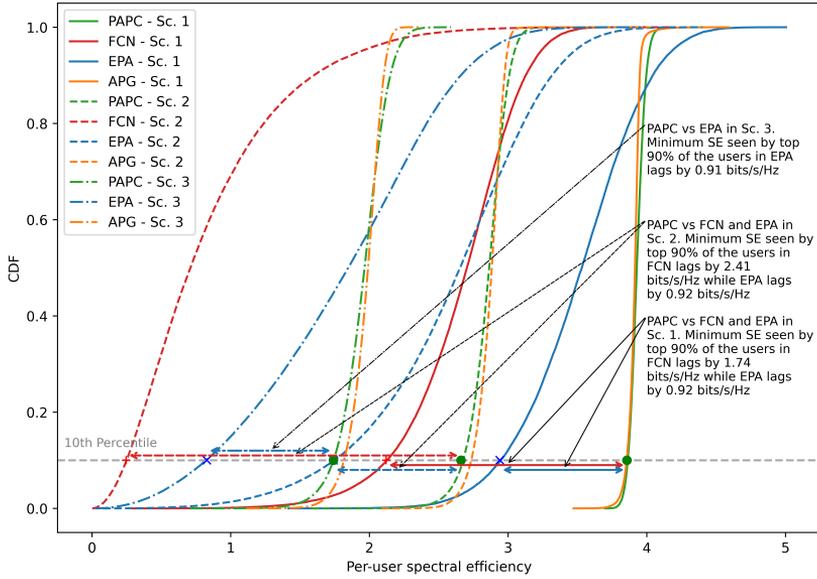


Figure 4.4. CDF comparison across Scenarios 1 to 3 for different algorithms. P consistently approaches APG performance, outperforming other algorithms due to its masking and attention mechanisms. FCN struggles due to its lack of structure and pilot allocation information.

| Algorithm | Run-time (in secs) |
|-----------|--------------------|
| APG | 38.7373 |
| P | 0.0262 |
| EPA | 0.0003 |

Table 4.2. Run-time of different algorithms in Scenario 3.

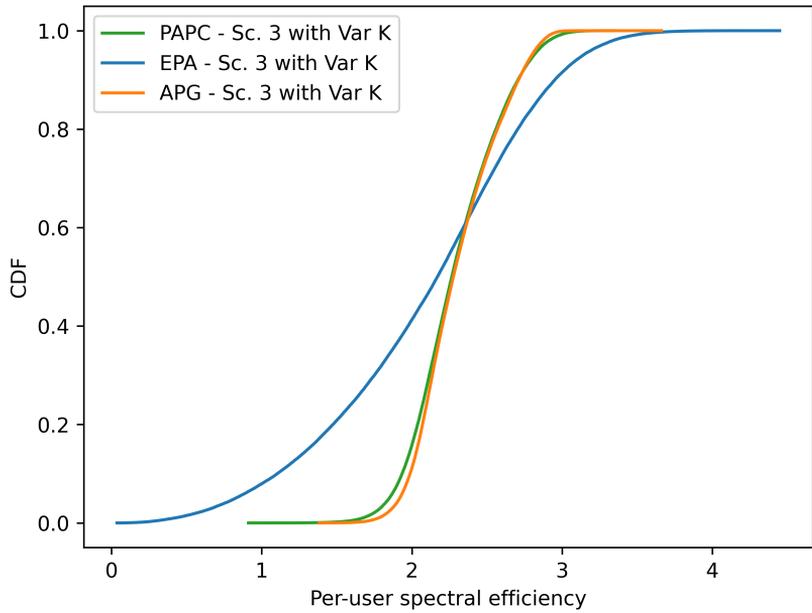


Figure 4.5. Run-time of the algorithms in Scenario 3. EPA is the fastest algorithm, while PACS achieves comparable performance to that of APG, but it is nearly 1000 times faster than APG.

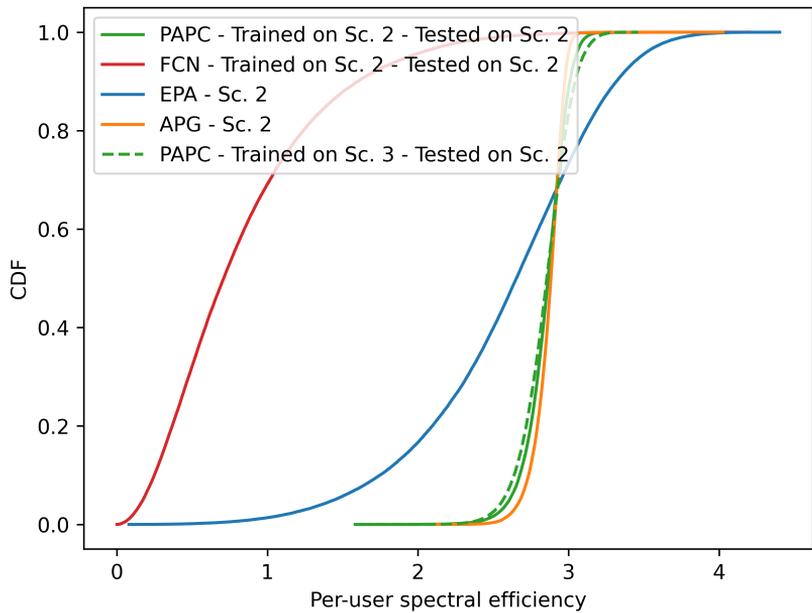


Figure 4.6. Comparison of PACSs trained on Scenario 2 and Scenario 3 (tested with $K = 40$), showing matching performance on Scenario 2 and validating that larger configurations with padding do not compromise results.

5. Conclusion and Future Work

5.1 Conclusion

This thesis has delved into a comprehensive exploration, analysis, and comparison of pilot resource allocation and power control techniques for massive MIMO systems. The research findings are encapsulated in four publications and provide an encompassing understanding of resource optimization within these systems. The significant insights and conclusions drawn from this investigation can be summarized as follows:

- Publications I and II have revealed the intricate relationship between pilot overhead, channel estimation, and SE in massive MIMO systems. The analytical expressions, derived in these studies, establish the critical importance of optimal pilot overhead allocation for enhancing performance in both uplink and downlink transmissions. The closed-form expressions serve as vital tools for selecting optimal pilot overhead parameters.
- The innovative ANN framework proposed in Publication III has significantly advanced the downlink power control in CFmMIMO systems. Achieving the performance close to the state-of-the-art methods, this methodology significantly curtails computational complexity.
- The PAPC architecture detailed in Publication IV further enhanced the downlink power control problem using additional custom designed architectural changes. With these advanced deep learning architectures, the PAPC enhanced the ANN to outperforms existing power control algorithms in computational efficiency while maintaining comparable performance. It presents a robust solution to the challenges associated with non-convexity and non-orthogonal pilots, inherent in power control.
- The comparative analysis of pilot resource allocation and power control tech-

niques highlighted in the thesis underscores the trade-offs among performance, complexity, and robustness. These insights are pivotal in guiding the design of efficient, practical algorithms for massive MIMO systems.

5.2 Future Work

Grounded on the insights obtained from this comprehensive study, several potential avenues for future research in the realm of resource optimization for massive MIMO systems emerge:

- Publications I and II in this thesis primarily concentrate on deriving analytical expressions pertaining to MRC beamforming. ZF beamforming, while more challenging, could potentially offer enhanced performance in massive MIMO systems. As such, a worthwhile avenue for future investigation might entail the derivation of similar analytical expressions for ZF beamforming.
- Another promising direction could be to rethink the performance metrics employed for designing pilot overhead. While our study uses average spectral efficiency as the metric, considering the dropout probability might provide a more realistic performance assessment. The development of analytical expressions for dropout probability, particularly with imperfect covariance estimates, presents an intriguing and challenging problem for future research.
- While the chosen method from [30] for covariance matrix estimation was justified for its tractability in deriving closed-form SE expressions, future work could explore alternative covariance estimation techniques. The results from such exploration could provide a richer basis for comparison and yield deeper insights into the trade-offs between estimation accuracy and pilot overhead. Additionally, such a comparison may further reinforce our choice of [30] as the most suitable method for evaluating the impact of the pilot overhead.
- The matter of pilot resource allocation in CFmMIMO systems is a crucial, yet complex, aspect that warrants further exploration. Moreover, the task of concurrently addressing pilot allocation and power control can elevate the challenges but is pivotal for maximizing the overall system performance. The development of innovative algorithms that tackle both aspects within a unified framework could be a fertile ground for research and significantly impact CFmMIMO systems.
- The power control techniques explored in Publications III and IV showed considerable promise. Further refinement and testing of these methods under varying scenarios, particularly those involving real-world constraints and non-

ideal conditions, could provide further valuable insights and developments.

References

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, November 2010.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, pp. 186–195, February 2014.
- [3] 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, pp. 742–758, Oct 2014.
- [4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5g be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1065–1082, June 2014.
- [5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, January 2013.
- [6] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 172–179, February 2013.
- [7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 160–171, February 2013.
- [8] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.
- [9] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, M. Debbah, and P. Xiao, "Max–min rate of cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. on Communications*, vol. 67, no. 10, pp. 6796–6815, 2019.
- [10] S. Chakraborty, O. T. Demir, E. Björnson, and P. Giselsson, "Efficient downlink power allocation algorithms for cell-free massive MIMO systems," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 168–186, 2021.
- [11] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2019.
- [12] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with mmse processing and centralized implementation," *IEEE Trans. on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2020.

References

- [13] O. T. Demir, E. Björnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO*. 2021.
- [14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proc. Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.
- [15] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [16] E. Nayebe, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, “Precoding and power optimization in cell-free massive MIMO systems,” *IEEE Trans. on Wireless Communications*, vol. 16, no. 7, pp. 4445–4459, 2017.
- [17] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, “Ubiquitous cell-free massive MIMO communications,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, pp. 1–13, Dec 2019.
- [18] A. Burr, M. Bashar, and D. Maryopi, “Cooperative access networks: Optimum fronthaul quantization in distributed massive MIMO and cloud RAN - invited paper,” in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, 2018.
- [19] H. Masoumi and M. J. Emadi, “Performance analysis of cell-free massive MIMO system with limited fronthaul capacity and hardware impairments,” *IEEE Trans. on Wireless Communications*, vol. 19, no. 2, pp. 1038–1053, 2020.
- [20] 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, pp. 2640–2651, August 2011.
- [21] A. Pitarokoilis, E. Björnson, and E. G. Larsson, “On the effect of imperfect timing synchronization on pilot contamination,” in *IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.
- [22] E. Björnson, L. Sanguinetti, and M. Debbah, “Massive MIMO with imperfect channel covariance information,” in *50th Asilomar Conference on Signals, Systems and Computers*, pp. 974–978, Nov 2016.
- [23] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [24] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO has unlimited capacity,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, 2018.
- [25] E. Björnson, J. Hoydis, and L. Sanguinetti, “Pilot contamination is not a fundamental asymptotic limitation in massive MIMO,” in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2017.
- [26] S. Haghhighatshoar and G. Caire, “Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation,” *IEEE Transactions on Wireless Communications*, vol. 16, pp. 8316–8332, Dec 2017.
- [27] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, “Sparse bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking,” *IEEE Transactions on Communications*, vol. 67, pp. 1925–1938, March 2019.
- [28] H. Xie, F. Gao, S. Zhang, and S. Jin, “A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3170–3184, 2017.

- [29] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Processing Letters*, vol. 25, pp. 863–867, June 2018.
- [30] K. Upadhyay and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Processing Letters*, vol. 25, pp. 546–550, April 2018.
- [31] M. Farooq, H. Q. Ngo, and L. N. Tran, "Accelerated projected gradient method for the optimization of cell-free massive MIMO downlink," in *Proc. IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK*, pp. 1–6, 2020.
- [32] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 25–39, 2018.
- [33] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1871–1874, 2017.
- [34] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive MIMO: A blessing in disguise," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5153–5169, 2019.
- [35] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 574–590, Jan 2018.
- [36] S. Ag, I. Viering, H. Hofstetter, W. Utschick, and I. Viering, "Spatial long-term variations in urban, rural and indoor environments," 2002.
- [37] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 142–148, 2014.
- [38] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Technical Report (TR) 38.901, 3rd Generation Partnership Project (3GPP), Dec. 2019. Version 16.1.0.
- [39] M. Farooq, H. Q. Ngo, E. Hong, and L. Tran, "Utility maximization for large-scale cell-free massive MIMO downlink," *IEEE Trans. on Communications*, vol. 69, no. 10, pp. 7050–7062, 2021.
- [40] N. Rajapaksha, K. B. Shashika Manosha, N. Rajatheva, and M. Latva-Aho, "Deep learning-based power control for cell-free massive MIMO networks," in *ICC 2021 - IEEE International Conference on Communications*, pp. 1–7, 2021.
- [41] N. Rajapaksha, K. B. S. Manosha, N. Rajatheva, and M. Latva-aho, "Unsupervised learning-based joint power control and fronthaul capacity allocation in cell-free massive MIMO with hardware impairments," *IEEE Wireless Communications Letters*, vol. 12, no. 7, pp. 1159–1163, 2023.
- [42] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised-learning power control for cell-free wireless systems," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–5, 2019.
- [43] A. Mazhari Saray and A. Ebrahimi, "Max-min power control of cell free massive MIMO system employing deep learning," in *2022 4th West Asian Symposium on Optical and Millimeter-wave Wireless Communications (WASOWC)*, pp. 1–4, 2022.
- [44] C. D'Andrea, A. Zappone, S. Buzzi, and M. Debbah, "Uplink power control in cell-free massive MIMO via deep learning," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 554–558, 2019.

- [45] Y. Zhang, J. Zhang, Y. Jin, S. Buzzi, and B. Ai, "Deep learning-based power control for uplink cell-free massive MIMO systems," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2021.
- [46] H. Bao, K. Xiong, R. Zhang, P. Fan, D. Niyato, and K. B. Letaief, "User cooperation with power control for federated learning in CFmMIMO networks," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 122–127, 2023.
- [47] I. M. Braga, R. P. Antonioli, G. Fodor, Y. C. B. Silva, and W. C. Freitas, "Decentralized joint pilot and data power control based on deep reinforcement learning for the uplink of cell-free systems," *IEEE Trans. on Vehicular Technology*, vol. 72, no. 1, pp. 957–972, 2023.
- [48] M. Rahmani, M. Bashar, M. J. Dehghani, P. Xiao, R. Tafazolli, and M. Debbah, "Deep reinforcement learning-based power allocation in uplink cell-free massive MIMO," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 459–464, 2022.
- [49] X. Zhang, M. Kaneko, V. An Le, and Y. Ji, "Deep reinforcement learning-based uplink power control in cell-free massive MIMO," in *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*, pp. 567–572, 2023.
- [50] D. Yu, H. Lee, S. Hong, and S. Park, "Learning decentralized power control in cell-free massive MIMO networks," *IEEE Trans. on Vehicular Technology*, vol. 72, no. 7, pp. 9653–9658, 2023.
- [51] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised-learning power allocation for the cell-free downlink," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–5, 2020.
- [52] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised learning for C-RAN power control and power allocation," *IEEE Communications Letters*, vol. 25, no. 3, pp. 687–691, 2021.
- [53] L. Luo, J. Zhang, S. Chen, X. Zhang, B. Ai, and D. W. K. Ng, "Downlink power control for cell-free massive MIMO with deep reinforcement learning," *IEEE Trans. on Vehicular Technology*, vol. 71, no. 6, pp. 6772–6777, 2022.
- [54] L. Salaün and H. Yang, "Deep learning based power control for cell-free massive MIMO with MRT," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 01–07, 2021.
- [55] C. Hao, T. T. Vu, H. Q. Ngo, M. N. Dao, X. Dang, C. Wang, and M. Matthaiou, "Joint user association and power control for cell-free massive MIMO," *IEEE Internet of Things Journal*, pp. 1–1, 2024.
- [56] L. Salaün, H. Yang, S. Mishra, and C. S. Chen, "A GNN approach for cell-free massive MIMO," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pp. 3053–3058, 2022.
- [57] M. Zaher, O. T. Demir, E. Björnson, and M. Petrova, "Learning-based downlink power allocation in cell-free massive MIMO systems," *IEEE Trans. on Wireless Communications*, vol. 22, no. 1, pp. 174–188, 2023.
- [58] Y. Zhang, J. Zhang, S. Buzzi, H. Xiao, and B. Ai, "Unsupervised deep learning for power control of cell-free massive MIMO systems," *IEEE Trans. on Vehicular Technology*, vol. 72, no. 7, pp. 9585–9590, 2023.
- [59] S. Chakraborty and B. R. Manoj, "Power allocation in a cell-free MIMO system using reinforcement learning-based approach," in *2023 National Conference on Communications (NCC)*, pp. 1–6, 2023.

Errata

Publication I

A. K. Kocharalakota, K. Upadhyaya and S. A. Vorobyov. On Achievable Rates for Massive MIMO System with Imperfect Channel Covariance Information. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, pp. 4504-4508, 2019.

© 2019

Reprinted with permission.

ON ACHIEVABLE RATES FOR MASSIVE MIMO SYSTEM WITH IMPERFECT CHANNEL COVARIANCE INFORMATION

Atchutaram K. Kocharalakota* Karthik Upadhya† Sergiy A. Vorobyov*

* Dept. of Signal Processing and Acoustics, Aalto University, Espoo, Finland

† Nokia Bell Labs, Espoo, Finland

Emails: kameswara.kocharalakota@aalto.fi, karthik.upadhya[at]nokia-bell-labs[dot]com, svor@ieee.org

ABSTRACT

An analytical lower bound on uplink channel capacity of a user in a massive multiple-input multiple-output system where the channel vector and the covariance matrices of the users in that cell are unknown is derived in this paper. This analytical bound enables us to choose appropriate sample size for covariance matrix estimation to meet the spectral efficiency requirements. The accurate agreement between the derived bound and the simulated bound based on random samples of channel vectors and covariance matrices is shown.

Index Terms— Achievable rate, spectral efficiency, Massive multiple-input multiple-output (MIMO), covariance estimation, channel estimation, pilot contamination.

1. INTRODUCTION

Increased spectral efficiency through spatial multiplexing makes massive multiple-input multiple-output (MIMO) system, where each base station (BS) has large number of antennas to serve multiple users within the cell, one of the key technologies for the next generation mobile networks [1–3]. However, knowledge of channel state information (CSI) at the BS is essential for the communication between the users and the BS.

A limited number of pilots are used for channel estimation due to finite coherence time and finite coherence bandwidth, which results in the pilot contamination problem in multi-cell scenarios [1]. It has been shown recently that, under certain assumptions on the spatial covariance matrices, the sum rate for the massive MIMO system is unbounded despite the presence of pilot contamination [4]. However, this result assumes availability of the individual user covariance matrices at the BS, which, in practice, are also contaminated. Several methods have been proposed in recent literature for estimating the spatial covariance matrices [5–8], and with the estimated channel and covariance information, bounds on the channel capacity are numerically studied in these papers.

However closed form expressions for such bounds utilizing estimated covariance matrices are not available in the

literature, to the best of our knowledge. Such bounds can provide useful insights into the number of observations needed for estimating the covariance matrices to achieve a target spectral efficiency (SE).

In this paper, we derive closed form expressions for the SE in a massive MIMO system with imperfect channel and covariance matrix estimates at the BS, thereby characterizing the SE in terms of the number of samples required to estimate the covariance matrices.

2. SYSTEM MODEL

A massive MIMO system with L cells having K users within each of them is considered. All the BSs are assumed to have M antennas each, and all the users have single antenna.

The uplink (UL) channel between the k^{th} user in the l^{th} cell, indexed as (l, k) , and a BS j is denoted as $\mathbf{h}_{lk} \in \mathbb{C}^M$ (the subscript j is dropped for the sake of simplicity), and is assumed to be a zero mean circularly symmetric complex Gaussian random vector, denoted as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{lk})$, where \mathbf{R}_{lk} is the covariance matrix. The channel is assumed to be constant for τ_c symbols—length of the coherence block, while its second order statistics are assumed to be constant for τ_s coherence blocks. C_u symbols are used for UL communication within each coherence block.

The UL received signal $\mathbf{Y} \in \mathbb{C}^{M \times C_u}$ in n^{th} coherence block at j^{th} BS, is given by:

$$\mathbf{Y}[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\mu} \mathbf{h}_{lk} \mathbf{x}_{lk}^T[n] + \mathbf{N}[n] \quad (1)$$

where μ is the transmit power at each user, $\mathbf{N} \in \mathbb{C}^{M \times C_u}$ is the additive white Gaussian noise whose elements are distributed as $\mathcal{CN}(0, 1)$, $\mathbf{x}_{lk} \in \mathbb{C}^{C_u}$ is the signal transmitted by user (l, k) whose elements are distributed as $\mathcal{CN}(0, 1)$. It contains the data signal as well as the pilot signals for estimating the channel and the covariance matrices.

In the following subsections, the pilot structures and estimation techniques for the channel vector and the covariance matrices are explained.

2.1. LMMSE Channel Estimation

The BSs and users are assumed to be perfectly synchronized, and K symbols in each coherence block are dedicated for UL channel estimation where each user in a cell is allocated a pilot from a set of K orthogonal sequences. Let $\mathbf{p}_k \in \mathbb{C}^K$, such that $\mathbf{p}_k^H \mathbf{p}_m = K\delta_{km}$, be the pilot transmitted by the k^{th} user in every cell, and $\mathbf{Y}^{(p)}[n] \in \mathbb{C}^{M \times K}$ be the received signal corresponding to pilot transmissions in the n^{th} coherence block.

The linear minimum mean squared error (LMMSE) estimate of the channel from the target user (j, u) in the n^{th} coherent block is given by:

$$\hat{\mathbf{h}}_{ju}[n] = \hat{\mathbf{R}}_{ju} \hat{\mathbf{Q}}_u^{-1} \hat{\mathbf{h}}_{ju}^{LS}[n], \quad n = 1, \dots, \tau_s \quad (2)$$

where $\hat{\mathbf{R}}_{ju}$ is the estimated covariance matrix, $\hat{\mathbf{Q}}_u$ is an estimate of \mathbf{Q}_u , and

$$\hat{\mathbf{h}}_{ju}^{LS}[n] = \frac{1}{K\sqrt{\mu}} \mathbf{Y}^{(p)}[n] \mathbf{p}_u^*, \quad (3)$$

$$\mathbf{Q}_u = \mathbb{E}\{\hat{\mathbf{h}}_{ju}^{LS} (\hat{\mathbf{h}}_{ju}^{LS})^H\} = \sum_{l=0}^{L-1} \mathbf{R}_{lu} + \frac{1}{K\mu} \mathbf{I}. \quad (4)$$

In the following subsection, we describe pilot structure and estimation techniques for these covariance matrices.

2.2. Covariance Matrix Estimation

Here, a covariance matrix estimation technique using the pilot structure introduced in [7] for estimating both $\hat{\mathbf{R}}_{ju}$ and $\hat{\mathbf{Q}}_u$ is described.

An additional set of pilot sequences $\{\phi_{lk}[n]\}_{n=1}^{N_R}$ for estimating \mathbf{R}_{ju} is transmitted by user (l, k). Then $\phi_{lk}[n]$ is the pilot sequence transmitted in n^{th} coherent block, and it is given by $\phi_{lk}[n] = [\bar{\phi}_k^T, e^{j\theta_{ln}} \bar{\phi}_k^T]^T \in \mathbb{C}^{C_r}$ ($2K \leq C_r < C_u$), and $\bar{\phi}_k$ is the sub-sequence used by k^{th} user in all the cells. It is infact a column chosen from Φ where $\Phi^H \Phi = K\mathbf{I}_{C_r/2}$. Here also $\{\theta_{ln}\}_{n=1}^{N_R}$ is the random phase sequence generated for all the users in l^{th} cell such that it is independent of the channel vectors [7]. Furthermore, it is chosen such that $\mathbb{E}(e^{j\theta_{ln}}) = 0$. These sequences are independently and identically distributed (i.i.d) over different cells, and are assumed to be known at corresponding BS and all the users in the cell.

Regularized covariance matrix \mathbf{R}_{ju} estimate and its expectation are given as follows:

$$\hat{\mathbf{R}}_{ju} \triangleq \alpha \ddot{\mathbf{R}}_{ju} + (1 - \alpha) \mathbf{R}_b, \quad (5)$$

where $\ddot{\mathbf{R}}_{ju} = \frac{1}{N_R} \sum_{n=1}^{N_R} \text{sym}(\hat{\mathbf{h}}_{ju}^{(1)}[n] (\hat{\mathbf{h}}_{ju}^{(2)}[n])^H)$, $\text{sym}(\cdot)$ is defined as $\text{sym}(\mathbf{A}) \triangleq \frac{1}{2}(\mathbf{A} + \mathbf{A}^H)$ and

$$\begin{aligned} \hat{\mathbf{h}}_{ju}^{(1)}[n] &= \mathbf{Y}^{(1)}[n] \sqrt{\mu} \bar{\phi}_u^* (\mu \bar{\phi}_u^T \bar{\phi}_u^*)^{-1} \\ &= \mathbf{h}_{ju} + \sum_{l \neq j} \mathbf{h}_{lu} + \frac{1}{K\sqrt{\mu}} \mathbf{N}^{(1)}[n] \bar{\phi}_u^*, \end{aligned} \quad (6)$$

$$\begin{aligned} \hat{\mathbf{h}}_{ju}^{(2)}[n] &= \mathbf{Y}^{(2)}[n] \sqrt{\mu} e^{-j\theta_{jn}} \bar{\phi}_u^* (\mu \bar{\phi}_u^T \bar{\phi}_u^*)^{-1} \\ &= \mathbf{h}_{ju} + \sum_{l \neq j} \mathbf{h}_{lu} e^{-j\theta_{jn}} + \frac{1}{K\sqrt{\mu}} \mathbf{N}^{(2)}[n] \bar{\phi}_u^* e^{-j\theta_{jn}}, \end{aligned} \quad (7)$$

$\mathbf{Y}^{(1)}[n]$ and $\mathbf{Y}^{(2)}[n]$ are the received signals that correspond to the first and second sub-sequences of ϕ , respectively, $\mathbf{N}^{(1)}[n]$ and $\mathbf{N}^{(2)}[n]$ are the noise signals that are additive to first and second pilot sub-sequence, respectively, \mathbf{R}_b is an arbitrary symmetric positive definite bias-matrix, and α is a design parameter. For later use, it is useful to define $\bar{\mathbf{R}}_{ju} \triangleq \mathbb{E}\{\hat{\mathbf{R}}_{ju}\} = \alpha \mathbf{R}_{ju} + (1 - \alpha) \mathbf{R}_b$.

For estimating \mathbf{Q}_u , an additional set of pilots is not required, but the received signal that corresponds to pilot sequence \mathbf{p}_u can be utilized to compute an unbiased estimator of \mathbf{Q}_u , which is given as follows:

$$\hat{\mathbf{Q}}_u = \frac{1}{N_Q} \sum_{n=1}^{N_Q} \hat{\mathbf{h}}_{ju}^{LS}[n] (\hat{\mathbf{h}}_{ju}^{LS}[n])^H. \quad (8)$$

In the following section, the SE for the UL channel of a single target user (j, u) is derived. For the derivation, we consider a matched filter receiver combiner, $\mathbf{v}_{ju}[n] = \hat{\mathbf{h}}_{ju}[n] = \hat{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{ju}^{LS}[n]$ where $\hat{\mathbf{W}}_{ju} \triangleq \hat{\mathbf{R}}_{ju} \hat{\mathbf{Q}}_u^{-1}$. It is assumed that $\hat{\mathbf{R}}_{ju}$, $\hat{\mathbf{Q}}_u$, and $\hat{\mathbf{h}}_{ju}^{LS}[n]$ are uncorrelated within a coherence block n , i.e., $\hat{\mathbf{R}}_{ju}$ and $\hat{\mathbf{Q}}_u$ are computed each from a different set of coherence blocks that does not include n . Furthermore, it is assumed that $N_Q > M$.

3. UL SPECTRAL EFFICIENCY

To obtain a lower bound on the channel capacity, we assume that the codeword is spread over multiple realizations of the covariance estimates. Then, a lower bound on capacity of the UL channel from user (j, u) to BS j is given by [6]:

$$R_{ju} = \left(1 - \frac{K}{C_u} - \frac{N_R C_r}{C_u \tau_s}\right) \log_2(1 + \gamma_u) \quad [\text{bits/s/Hz}] \quad (9)$$

where γ_u is given in (10) at the top of the next page and $\mathbf{R}_s \triangleq \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} \mathbf{R}_{lk} + \frac{1}{\mu} \mathbf{I}$. The expectation taken in all the terms is over the random matrix $\hat{\mathbf{W}}_{ju}$.

Before deriving the expectation terms of (10), we give lemmas that will be useful in the derivation. In what follows, \mathbb{E}_R represents the expectation over $\hat{\mathbf{R}}_{ju}$, \mathbb{E}_Q represents expectation over $\hat{\mathbf{Q}}_u$, and \mathbb{E} represents expectation over both.

Lemma 1. Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$, and for any mutually independent M -dimensional random vectors \mathbf{h}_1 , \mathbf{h}_2 , and \mathbf{h} distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_1)$, $\mathcal{CN}(\mathbf{0}, \mathbf{R}_2)$, and $\mathcal{CN}(\mathbf{0}, \mathbf{R})$, respectively, we have

$$\mathbb{E}\{\mathbf{h}_1 \mathbf{h}_2^H \mathbf{A} \mathbf{h}_2 \mathbf{h}_1^H\} = \mathbf{R}_1 \text{tr}(\mathbf{A} \mathbf{R}_2), \quad (13)$$

$$\mathbb{E}\{\mathbf{h} \mathbf{h}^H \mathbf{A} \mathbf{h} \mathbf{h}^H\} = \mathbf{R} \mathbf{A} \mathbf{R} + \mathbf{R} \text{tr}(\mathbf{A} \mathbf{R}). \quad (14)$$

$$\gamma_u = \frac{|\mathbb{E}\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{ju})\}|^2}{\mathbb{E}\{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_u \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s)\} + \sum_{l=1}^L \mathbb{E}\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{lu})|^2\} - |\mathbb{E}\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{ju})\}|^2} \quad (10)$$

$$\mathbb{E}\{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_u \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s)\} = \kappa_1 \text{tr}(\bar{\mathbf{W}}_{ju} \mathbf{Q}_u \bar{\mathbf{W}}_{ju}^H \mathbf{R}_s) + \frac{\alpha^2 \kappa_1}{2N_R} \left\{ M \text{tr}(\mathbf{R}_s \mathbf{Q}_u) + \text{tr}(\mathbf{W}_{ju}) \text{tr}(\mathbf{R}_s \mathbf{R}_{ju}) \right\} \quad (11)$$

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{lu})|^2\} &= \kappa_2 |\text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{lu})|^2 + \frac{\alpha^2 \kappa_2}{2N_R} \text{tr}(\mathbf{W}_{lu} \mathbf{Q}_u \mathbf{W}_{lu}^H \mathbf{Q}_u) + \frac{\alpha^2 \kappa_2}{2N_R} \text{tr}(\mathbf{W}_{lu} \mathbf{R}_{ju} \mathbf{W}_{lu}^H \mathbf{R}_{ju}) \\ &+ \frac{\kappa_1}{N_Q} \text{tr}(\bar{\mathbf{W}}_{ju}^2 \mathbf{Q}_u \mathbf{W}_{lu}^2 \mathbf{Q}_u) + \frac{\alpha^2 \kappa_1}{2N_Q N_R} M \text{tr}(\mathbf{W}_{lu}^2 \mathbf{Q}_u^2) + \frac{\alpha^2 \kappa_1}{2N_Q N_R} \text{tr}(\mathbf{W}_{ju}) \text{tr}(\mathbf{W}_{lu}^2 \mathbf{Q}_u \mathbf{R}_{ju}) \end{aligned} \quad (12)$$

Proof. (13) can be derived by splitting the expectation over \mathbf{h}_1 and \mathbf{h}_2 . Proof of (14) is straightforward, but it involves the values of second and fourth order moments of Gaussian random variables. Thus, it is left to the reader because of space limitation. \square

Lemma 2. Given a Hermitian matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$, an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$, and a complex Wishart matrix, $\mathbf{X} \in \mathbb{C}^{M \times M}$, with N degrees of freedom (represented as $\mathcal{CW}(N, \mathbf{I})$), we have

$$\mathbb{E}\{[\mathbf{X}^{-1}]_{ij}\} = \frac{[\mathbf{I}]_{ij}}{N-M}, \quad (15)$$

$$\mathbb{E}\{[\mathbf{X}^{-1}]_{ij} [\mathbf{X}^{-1}]_{lk}\} = \frac{[\mathbf{I}]_{ij} [\mathbf{I}]_{lk} + \frac{1}{N-M} [\mathbf{I}]_{lj} [\mathbf{I}]_{ik}}{(N-M)^2 - 1}, \quad (16)$$

$$\mathbb{E}\{\text{tr}(\mathbf{X}^{-2} \mathbf{C})\} = \frac{N}{(N-M)^3 - (N-M)} \text{tr}(\mathbf{C}), \quad (17)$$

$$\mathbb{E}\{|\text{tr}(\mathbf{X}^{-1} \mathbf{A})|^2\} = \frac{|\text{tr}(\mathbf{A})|^2 + \frac{1}{N-M} \text{tr}(\mathbf{A} \mathbf{A}^H)}{(N-M)^2 - 1}. \quad (18)$$

Proof. Proof is available in Appendix. \square

Lemma 3. Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$, we have

$$\begin{aligned} \mathbb{E}\{\ddot{\mathbf{R}}_{ju} \mathbf{A} \ddot{\mathbf{R}}_{ju}\} &= \mathbf{R}_{ju} \mathbf{A} \mathbf{R}_{ju} + \frac{1}{2N_R} \mathbf{Q}_u \text{tr}(\mathbf{A} \mathbf{Q}_u) \\ &+ \frac{1}{2N_R} \mathbf{R}_{ju} \text{tr}(\mathbf{A} \mathbf{R}_{ju}) \end{aligned} \quad (19)$$

and

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\ddot{\mathbf{R}}_{ju} \mathbf{A})|^2\} &= |\text{tr}(\mathbf{R}_{ju} \mathbf{A})|^2 + \frac{1}{2N_R} \text{tr}(\mathbf{A} \mathbf{Q}_u \mathbf{A}^H \mathbf{Q}_u) \\ &+ \frac{1}{2N_R} \text{tr}(\mathbf{A} \mathbf{R}_{ju} \mathbf{A}^H \mathbf{R}_{ju}) \end{aligned} \quad (20)$$

Proof. Proof of this lemma uses Lemma 1 and is presented in Appendix. \square

Now we are ready to formulate the key theorem.

Theorem 1. The signal component of (10) is given by

$$\mathbb{E} \text{tr}\{\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{ju}\} = \frac{N_Q}{N_Q - M} \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{ju}). \quad (21)$$

The first and second terms of the denominator in (10) are given in (11) and (12) at the top of this page, where $\kappa_1 = N_Q \kappa_2 / (N_Q - M)$, $\kappa_2 = N_Q^2 / ((N_Q - M)^2 - 1)$, $\bar{\mathbf{W}}_{ju} \triangleq \bar{\mathbf{R}}_{ju} \mathbf{Q}_u^{-1}$, and $\mathbf{W}_{lu} = \mathbf{R}_{lu} \mathbf{Q}_u^{-1}$.

Proof. We define a matrix $\tilde{\mathbf{Q}}_{ju}$ as follows:

$$\tilde{\mathbf{Q}}_{ju} \triangleq N_Q (\mathbf{Q}_{ju}^{-\frac{1}{2}} \hat{\mathbf{Q}}_u \mathbf{Q}_{ju}^{-\frac{1}{2}}). \quad (22)$$

It can be seen that $\tilde{\mathbf{Q}}_{ju}$ is a Wishart matrix distributed as $\mathcal{W}(N_Q, \mathbf{I})$. Using $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{ju} \hat{\mathbf{Q}}_u^{-1}$ and (22), the numerator term of (10) can be written as:

$$\mathbb{E} \text{tr}\{\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{ju}\} = N_Q \mathbb{E} \text{tr}\{\mathbf{Q}_{ju}^{-\frac{1}{2}} \tilde{\mathbf{Q}}_{ju}^{-1} \mathbf{Q}_{ju}^{-\frac{1}{2}} \hat{\mathbf{R}}_{ju} \mathbf{R}_{ju}\}. \quad (23)$$

By taking direct expectation over $\hat{\mathbf{R}}_{ju}$ in (23) and also using Lemma 2, (21) can be obtained.

Proof of (11) and (12) is as follows. By substituting $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{ju} \hat{\mathbf{Q}}_u^{-1}$ and (22) into the first and second denominator terms of (10) and by using Lemma 2, we get the following equations

$$\mathbb{E} \text{tr}\{\hat{\mathbf{W}}_{ju} \mathbf{Q}_u \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s\} = \kappa_1 \mathbb{E}_{R_R} \text{tr}\{\mathbf{Q}_u^{-1} \hat{\mathbf{R}}_{ju} \mathbf{R}_s \hat{\mathbf{R}}_{ju}\}, \quad (24)$$

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{lu})|^2\} &= \kappa_2 \mathbb{E}_{R_R} \{|\text{tr}(\mathbf{Q}_u^{-1} \hat{\mathbf{R}}_{ju} \mathbf{R}_{lu})|^2\} \\ &+ \frac{\kappa_1}{N_Q} \mathbb{E}_{R_R} \text{tr}\{\mathbf{Q}_u^{-1} \hat{\mathbf{R}}_{ju} \mathbf{R}_{lu} \mathbf{Q}_u^{-1} \mathbf{R}_{lu} \hat{\mathbf{R}}_{ju}\} \end{aligned} \quad (25)$$

By using Lemma 3, and by substituting (5) into (24) and (25), we get (11) and (12), respectively. \square

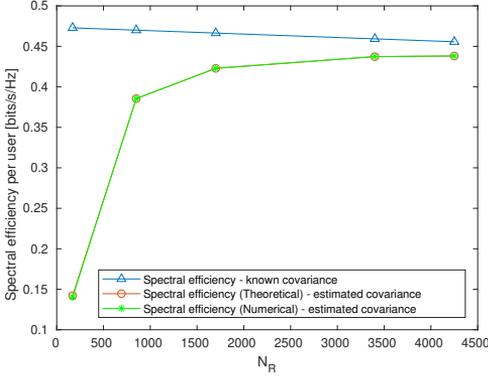


Fig. 1. UL SE of a user in massive MIMO system

4. SIMULATIONS

We consider a massive MIMO system with $L = 7$ cells each having $K = 10$ users, and the number of antennas at the BS is $M = 100$. The BSs are separated $300m$ and the users are uniformly located in a circle of radius $120m$ from the BS. The signal to noise ratio of the received signal from a user that is at a distance d is given by $78.6 - 37.6 \log_{10} d$. In Fig 1, we compare SE of a user in the central cell for the case of known covariance matrices, theoretical lower bound on channel capacity for the estimated covariance case and simulated SE for the case of estimated covariance matrices.

Number of UL resources used in a coherence block is chosen to be $C_u = 100$ symbols, and second order statistics are assumed to be constant for $\tau_s = 25000$ coherence blocks. The transmit power of the target is $\mu = 1$. Additionally, we choose $C_r = 2K$ symbols, $N_Q = N_R (> M)$, $\alpha = 0.95$, and $\mathbf{R}_b = \mathbf{I}$. Sample averaging for all the expectation terms is computed for 500 iterations for different values of $N_R = (170, 850, 1700, 3400, 4250)$.

It can be seen from Fig 1 that the theoretical achievable rate for the case of unknown covariance matrix asymptotically approaches the SE for the known covariance case. Also, the simulated SE matches the theoretical values.

5. CONCLUSION

An analytical expression for the SE of a user in a massive MIMO system is derived for the case when the matched filter receiver combiner uses estimated channel covariance matrices and estimated channel vector. The simulation results matched that obtained by the theoretical expression derived in this paper. The accurate agreement between the derived analytical bound for the SE and the results of simulations is demonstrated.

6. APPENDIX

Proof of Lemma 2. Proofs of (15) and (16) are given in [9].

Using eigen value decomposition $\mathbf{C} = \mathbf{U}\mathbf{A}\mathbf{U}^H$ and denoting $\tilde{\mathbf{X}} = \mathbf{U}^H \mathbf{X} \mathbf{U} \sim \mathcal{W}(N, \mathbf{I})$, (17) can be proved as follows:

$$\begin{aligned} \mathbb{E}\{\text{tr}\{\mathbf{X}^{-2}\mathbf{C}\}\} &= \mathbb{E}\{\text{tr}\{\tilde{\mathbf{X}}^{-2}\mathbf{A}\}\} = \sum_{i=1}^M [\mathbb{E}\{\tilde{\mathbf{X}}^{-2}\}]_{ii} [\mathbf{A}]_{ii} \\ &= [\mathbb{E}\{\tilde{\mathbf{X}}^{-2}\}]_{11} \text{tr}(\mathbf{A}) = \sum_{j=1}^M \mathbb{E}\{[\tilde{\mathbf{X}}^{-1}]_{1j} [\tilde{\mathbf{X}}^{-1}]_{j1}\} \text{tr}(\mathbf{C}) \\ &= \frac{N}{(N-M)^3 - (N-M)} \text{tr}(\mathbf{C}). \end{aligned}$$

The above derivation uses the fact that $[\mathbb{E}\{\tilde{\mathbf{X}}^{-2}\}]_{ii}$ is same for all $i \in \{1 \dots M\}$.

For (18), $\mathbb{E}\{|\text{tr}(\mathbf{X}^{-1}\mathbf{A})|^2\}$ is expanded as follows:

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\mathbf{X}^{-1}\mathbf{A})|^2\} &= \sum_{p,q,r,s=1}^M \mathbb{E}\{[\mathbf{X}^{-1}]_{pq} [\mathbf{X}^{-1}]_{sr}\} [\mathbf{A}]_{qp} [\mathbf{A}^H]_{rs} \\ &= \sum_{p=1}^M \sum_{s=1}^M \mathbb{E}\{[\mathbf{X}^{-1}]_{pp} [\mathbf{X}^{-1}]_{ss}\} [\mathbf{A}]_{pp} [\mathbf{A}^H]_{ss} \\ &+ \sum_{p=1}^M \sum_{s=1}^M \mathbb{E}\{[\mathbf{X}^{-1}]_{ps} [\mathbf{X}^{-1}]_{sp}\} [\mathbf{A}]_{sp} [\mathbf{A}^H]_{ps}. \end{aligned}$$

Using (16), the above equation can be re-written as in (18). \square

Proof of Lemma 3. Let us define a pair of mutually independent random vectors $\mathbf{g}_{jju}^{(1)}[n] \triangleq \hat{\mathbf{h}}_{jju}^{(1)}[n] - \mathbf{h}_{ju}$ and $\mathbf{g}_{jju}^{(2)}[n] \triangleq \hat{\mathbf{h}}_{jju}^{(2)}[n] - \mathbf{h}_{ju}$. Their covariance matrices are identically equal to $\mathbf{Q}_u - \mathbf{R}_{ju}$. Additionally, we also define mutually independent set of matrices as $\check{\mathbf{R}}_{ju}[n] \triangleq \text{sym}(\hat{\mathbf{h}}_{ju}^{(1)}[n] (\hat{\mathbf{h}}_{ju}^{(2)}[n])^H)$, $\forall n \in \{1, \dots, N_R\}$ such that $\check{\mathbf{R}}_{ju} = \frac{1}{N_R} \sum_{n=1}^{N_R} \check{\mathbf{R}}_{ju}[n]$.

Using the definitions of $\mathbf{g}_{jju}^{(1)}[n]$ and $\mathbf{g}_{jju}^{(2)}[n]$, and also Lemma 1, it can be shown that

$$\begin{aligned} \mathbb{E}\{\check{\mathbf{R}}_{ju}[n] \mathbf{A} \check{\mathbf{R}}_{ju}[n]\} &= \mathbf{R}_{ju} \mathbf{A} \mathbf{R}_{ju} + \frac{1}{2} \mathbf{Q}_u \text{tr}(\mathbf{A} \mathbf{Q}_u) \\ &+ \frac{1}{2} \mathbf{R}_{ju} \text{tr}(\mathbf{A} \mathbf{R}_{ju}), \quad \forall n = 1 \text{ to } N_R, \end{aligned} \quad (26)$$

and

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\check{\mathbf{R}}_{ju}[n] \mathbf{A})|^2\} &= |\text{tr}(\mathbf{R}_{ju} \mathbf{A})|^2 + \frac{1}{2} \text{tr}(\mathbf{A} \mathbf{Q}_u \mathbf{A}^H \mathbf{Q}_u) \\ &+ \frac{1}{2} \text{tr}(\mathbf{A} \mathbf{R}_{ju} \mathbf{A}^H \mathbf{R}_{ju}), \quad \forall n = 1 \text{ to } N_R. \end{aligned} \quad (27)$$

Finally, along with the equation $\check{\mathbf{R}}_{ju} = \frac{1}{N_R} \sum_{n=1}^{N_R} \check{\mathbf{R}}_{ju}[n]$, (26) and (27) will result in (19) and (20), respectively. \square

7. REFERENCES

- [1] 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.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
- [3] 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, Oct 2014.
- [4] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive mimo has unlimited capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, Jan 2018.
- [5] S. Haghghatshoar and G. Caire, "Massive mimo pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8316–8332, Dec 2017.
- [6] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive mimo with imperfect channel covariance information," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 974–978.
- [7] K. Upadhyay and S. A. Vorobyov, "Covariance matrix estimation for massive mimo," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 546–550, April 2018.
- [8] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive mimo," *IEEE Signal Processing Letters*, vol. 25, no. 6, pp. 863–867, June 2018.
- [9] D. Maiwald and D. Kraus, "On moments of complex wishart and complex inverse wishart distributed matrices," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1997, vol. 5, pp. 3817–3820 vol.5.

Publication II

A. K. Kocharlakota, K. Upadhyaya and S. A. Vorobyov. Impact of Pilot Overhead and Channel Estimation on the Performance of Massive MIMO. *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8242-8255, Dec 2021.

© 2021

Reprinted with permission.

Impact of Pilot Overhead and Channel Estimation on the Performance of Massive MIMO

Atchutaram K. Kocharalakota¹, *Student Member, IEEE*, Karthik Upadhyya, *Member, IEEE*,
and Sergiy A. Vorobyov², *Fellow, IEEE*

Abstract—This paper studies the impact of additional pilot overhead for covariance matrix estimation in a time-division duplexed (TDD) massive multiple-input multiple-output (MIMO) system. We choose average uplink (UL) and downlink (DL) spectral efficiencies (SEs) as performance metrics for the massive MIMO system, and derive closed form expressions for them in terms of the additional pilot overhead. The expressions are derived by considering linear minimum mean squared error (LMMSE)-type and element-wise LMMSE-type channel estimates that represent LMMSE and element-wise LMMSE with estimated covariance matrices, respectively. Using these expressions, a detailed theoretical analysis of SE behavior as a function of pilot overhead for both LMMSE-type and element-wise LMMSE-type channel estimation are presented, followed by simulations, which also demonstrate and validate theoretical results.

Index Terms—Spectral efficiency, massive multiple-input multiple-output (MIMO), covariance estimation, channel estimation, pilot contamination.

I. INTRODUCTION

A MULTI-USER massive multiple-input multiple-output (MIMO) system comprises multiple cells, each having a base station (BS) with a large number of antennas (hundreds) to serve multiple users (tens) within the cell. It is considered to be one of the key technologies for the fifth-generation (5G) cellular systems due to the considerable improvement in spectral efficiency (SE) through spatial multiplexing [1]–[5] achieved with low computational complexity [1], [6], [7]. However, acquiring channel state information (CSI) at the base station (BS) is essential to realize the benefits of a massive MIMO system.

We consider a time-division duplexing (TDD) massive MIMO system where the CSI is acquired through uplink (UL) pilots. In time-variant channels, the channels in two different coherence blocks, which is a collection of symbols within a coherence time and bandwidth, are uncorrelated. Consequently, the channel has to be estimated in each coherence

block. The number of orthogonal pilots available for channel estimation in a coherence block is limited by the number of available symbols in the coherence block that are not reserved for UL data and DL data, and as a result, UL pilot sequences need to be reused by users across the cells, causing the pilot contamination problem [1], [8], [9].

Despite the presence of pilot contamination, under the assumption that the covariance matrices of interfering users are asymptotically linearly independent to each other, the sum rate of the massive MIMO system has been proven to be unbounded [10]. However, the authors assume that contamination-free covariance matrices of individual users are available at the BS, while, in practice, these covariance matrices also have to be estimated at the BS. Covariance matrix estimation in a multi-cell TDD massive MIMO system is a non-trivial task because the channel estimates from which the covariance matrix estimates are obtained are themselves contaminated. Naively utilizing the contaminated channel estimates in a sample covariance estimator will result in the target user covariance matrix estimate containing the covariance matrices of the interference users. The algorithm that estimates the target covariance matrix in such a setup needs additional information from the users to isolate the target user covariance from the contaminated covariance; this is typically done using additional pilots.

Methods for estimating the individual covariance matrices in the presence of pilot contamination have been recently studied in [11]–[16]. In all these works, the authors assume that the channel covariance matrices are constant across multiple coherence blocks, and then, the observations from a few of these coherence blocks are used to estimate the covariance matrices. In [11], the authors first estimate the angle-delay power spread function from the contaminated channel estimates of multiple coherence blocks, then use this function for supervised/unsupervised clustering of the multipath components belonging to the target user. Finally, they use the clusters to estimate the spatial covariance matrix of the target user. In [14], the authors develop a method where the pilot allocation is changed in each coherence block. The channel estimates obtained from these blocks are then used to obtain a maximum-likelihood estimate of the contamination-free covariance matrix. Work [15] presents two methods which avoid contamination in the covariance matrices by utilizing dedicated orthogonal pilots for each user for estimating its individual spatial covariance matrix. In [16], a new pilot structure and a covariance matrix estimation method are developed that offer higher throughput and lower mean squared

Manuscript received March 12, 2021; revised August 1, 2021; accepted September 1, 2021. Date of publication September 13, 2021; date of current version December 16, 2021. This work was supported by the Academy of Finland under grant 319822. The associate editor coordinating the review of this article and approving it for publication was A. García Armada. (*Corresponding author: Sergiy A. Vorobyov.*)

Atchutaram K. Kocharalakota and Sergiy A. Vorobyov are with the Department of Signal Processing and Acoustics, Aalto University, 00076 Aalto, Finland (e-mail: kameswara.kocharalakota@aalto.fi; svor@ieee.org).

Karthik Upadhyya was with the Department of Signal Processing and Acoustics, Aalto University, 00076 Aalto, Finland. He is now with Nokia Bell Labs, 02610 Espoo, Finland (e-mail: karthik.upadhyya@gmail.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCOMM.2021.3112213>.

Digital Object Identifier 10.1109/TCOMM.2021.3112213

error (MSE) of the channel estimates compared to the method in [15]. Although [16] requires additional pilots for estimating the individual covariance matrices of each user, it does not assume any additional structures on the covariance matrices of the users unlike [11]–[13], and it does not require backhaul communication between the neighboring cells unlike [14]. Moreover, since the additional pilots in [16] are not dedicated to each user as in [15], the number of additional pilots in [16] does not grow with the total number of users in the entire system. Therefore, in this paper, we choose [16] to study the performance of covariance estimation method in a massive MIMO system. In particular, we emphasize the impact of pilot overhead or choice of channel estimation method on the performance of a massive MIMO system.

Utilization of the estimated covariance matrices for channel estimation results in a trade-off for the system performance. Indeed, increase of the number of additional pilots for estimating the covariance matrices will not only improve the quality of the covariance estimate (and hence, the channel estimate) but also increase the pilot overhead. Consequently, choosing the additional pilot overhead related to estimating the covariance matrices becomes a key trade-off problem for the system performance.

Except [11], in all the covariance estimation papers mentioned above, the authors derive SE expressions corresponding to a single realization of covariance estimate. Such an SE is achievable for a practical receiver which does not have perfect covariance information. However, one can notice, these papers use a numerically computed average SE as a performance metric of covariance estimation method in a massive MIMO system. Therefore, in this paper, we utilize the SE value that is averaged over multiple realizations (ensemble average) of the covariance estimates as a performance metric of the covariance estimation method in [16]. We first derive average SE expressions for two types of channel estimation methods namely: (1) LMMSE-type and (2) element-wise LMMSE-type channel estimation methods, that use estimated covariance matrices.¹ Note that, in this paper, we use *LMMSE-type/element-wise LMMSE-type* to denote the channel estimation with estimated covariance matrices, and *LMMSE/element-wise LMMSE* to denote channel estimation with true covariance matrices. Using the derived expressions, we demonstrate the impact of pilot overhead and channel estimation method on the performance of the massive MIMO system with imperfect covariance information at BSs.

The following are the contributions of this paper.

- We first derive closed-form expressions for the average UL and DL spectral efficiencies when the LMMSE-type and element-wise LMMSE-type channel estimates are used in a maximum ratio combiner (MRC) UL receiver, and in a matched filter precoding DL transmitter.
- Using theoretical and simulation studies on the derived SE expressions, we establish the fact that the number of additional pilots for covariance estimation is a key trade-off parameter that needs to be chosen optimally.

- Using these expressions, we then quantitatively compare the performance of the element-wise LMMSE-type channel estimate with the LMMSE-type channel estimate. To the best of our knowledge, this is the first work that quantitatively compares the average UL/DL SE obtained with LMMSE-type and element-wise LMMSE-type estimates.

The paper is organized as follows. In Section II, we describe the system model along with a brief review of the channel estimation and covariance matrices estimation methods we want to study. Section III reports the derivations of the closed-form expressions for the UL and DL SEs for both the channel estimation techniques described above. In Section IV, we present a detailed theoretical discussion on the impact of pilot overhead and channel estimation technique on massive MIMO. In Section V we provide the simulation results that demonstrate the theoretical conclusions made in Section IV. We conclude this work in Section VI. Technical proofs of lemmas and theorems in the paper appear in appendices at the end of the paper.

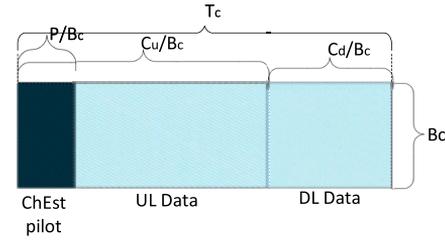
Notation: We use boldface capital letters for matrices, and boldface lowercase letters for vectors. The superscripts $(\cdot)^*$, $(\cdot)^\top$, and $(\cdot)^H$ denote element-wise conjugate, transpose, and Hermitian transpose operations, respectively. Moreover, $\mathcal{CN}(\mathbf{m}, \mathbf{R})$ denotes (circularly symmetric) complex Gaussian random vector with mean vector \mathbf{m} and covariance matrix \mathbf{R} , while $\mathcal{W}(N, \mathbf{R})$ denotes Wishart random matrix with N degrees of freedom and \mathbf{R} is the covariance matrix that corresponds to underlying Gaussian random vectors. In addition, $\mathcal{U}[x_1, x_2]$ stands for the uniform distribution between x_1 and x_2 . The element in i^{th} row and j^{th} column of the matrix \mathbf{A} is denoted as $[\mathbf{A}]_{ij}$, \mathbf{I} stands for an identity matrix (of appropriate size), $\text{diag}(\mathbf{A})$ is a diagonal matrix whose diagonal elements are same as the diagonal elements of the matrix \mathbf{A} . We use $\text{tr}(\cdot)$ to denote trace of a matrix, $\|\cdot\|$ to denote l_2 norm of a vector or a matrix, i.e., Frobenius norm, and $\mathbb{E}\{\cdot\}$ stands for the mathematical expectation. Finally, the symbol δ_{ij} is the Kronecker delta such that $\delta_{ij} = 1$ if $i = j$, and 0 otherwise.

II. SYSTEM MODEL

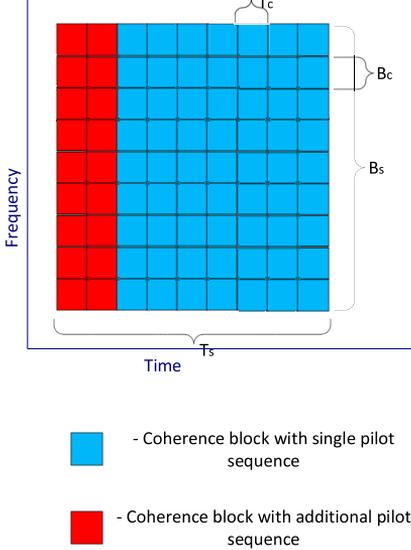
We consider a massive MIMO system with L cells, each having a BS with M antennas and serving K single-antenna users. We make a realistic assumption that the channels between users and BSs are spatially correlated [18]. The channel between user (l, k) (k^{th} user in l^{th} cell) and BS in the j^{th} cell is denoted as $\mathbf{h}_{jlk} \in \mathbb{C}^M$ and is assumed to be distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{jlk})$, where $\mathbf{R}_{jlk} \triangleq \mathbb{E}\{\mathbf{h}_{jlk}\mathbf{h}_{jlk}^H\}$ is the spatial covariance matrix. We consider the block-fading model where the channel is assumed to be constant over the coherence bandwidth B_c and coherence time T_c . In other words, the channel is assumed to be constant over a coherence block containing $\tau_c = B_c T_c$ symbols.

We consider TDD transmission and each coherence block is divided into slots for UL pilots, UL and DL data. The number of data symbols in the UL and DL time slot is denoted as C_u and C_d , respectively. The channel is assumed to be reciprocal, i.e., the DL channel between BS j and user (l, k) can be written as \mathbf{h}_{jlk}^* . This is represented in Fig. 1(a).

¹Some preliminary results are also reported in [17].



(a) Coherence block with only ChEst pilots.



(b) Grid of coherence blocks with coherent covariance matrices.

Fig. 1. Time frequency grid and pilot positioning.

We consider two types of UL pilots, namely, (i) pilots for estimating the channel (also referred to as ChEst pilots) and (ii) pilots for estimating the covariance matrix (referred to as CovEst pilots). Both ChEst pilots and CovEst pilots are assumed to be of length P symbols.

The spatial covariance matrices are assumed to be constant over a considerably longer time-interval and bandwidth than a single coherence block [11], [14]–[16], [19].² Specifically, we assume that the covariance matrices are coherent over the time-interval T_s and system bandwidth B_s , which implies that they can be assumed to be constant over $\tau_s = B_s T_s / \tau_c$ coherence blocks (usually several tens of thousands of blocks in practice). This time-frequency grid over which the second-order statistics of the channel are assumed to be constant is illustrated in Fig. 1(b).

Each of the τ_s coherence blocks contain ChEst pilots for channel estimation, whereas only N_F out of the τ_s coherence

²Note that, according to [19], this assumption is valid of urban and rural environment. However, this is questionable for indoor scenarios.

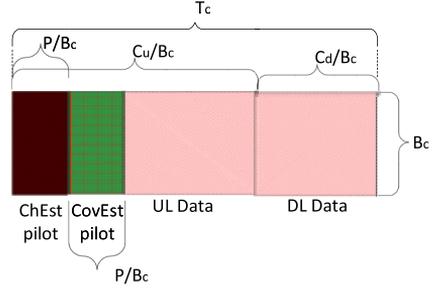


Fig. 2. Coherence block with additional CovEst pilots.

blocks contain CovEst pilots in addition to the ChEst pilots (as can be seen in Fig. 1(b)). The coherence blocks that contain the CovEst pilots are depicted in Fig. 2.

The UL received signal, $\mathbf{Y}_j[n] \in \mathbb{C}^{M \times C_u}$, in the n^{th} coherence block at BS j is given as

$$\mathbf{Y}_j[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\mu} \mathbf{h}_{jlk} \mathbf{x}_{lk}^T[n] + \mathbf{N}_j[n] \quad (1)$$

where $\mathbf{x}_{lk} \in \mathbb{C}^{C_u}$ is the signal transmitted by user (l, k) , $\mathbf{N}_j \in \mathbb{C}^{M \times C_u}$ is the additive white Gaussian noise (AWGN) at the BS, and μ is the UL transmit power. The transmitted data \mathbf{x}_{lk} is assumed to be distributed as $\mathbf{x}_{lk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ whereas the elements of \mathbf{N}_j are assumed to be identically and independently distributed (i.i.d) as $\mathcal{CN}(0, 1)$.

In the DL, the received signal $\mathbf{z}_{ju}[n] \in \mathbb{C}^{C_d}$ at user (j, u) in the n^{th} coherence block can be written as

$$\mathbf{z}_{ju}[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\lambda} (\mathbf{h}_{jlu}^H \mathbf{b}_{lk}) \mathbf{d}_{lk}[n] + \mathbf{e}[n]$$

where $\mathbf{d}_{lk} \in \mathbb{C}^{C_d}$ is the payload data from BS l to its user (l, k) , $\mathbf{b}_{lk} \in \mathbb{C}^M$ is the corresponding precoding vector normalized such that the average transmitted power is λ , i.e., $\mathbb{E}\{\|\mathbf{b}_{lk}\|^2\} = 1$, and $\mathbf{e} \in \mathbb{C}^{C_d}$ is the AWGN noise distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I})$.

A. Channel Estimation

A dedicated set of P ($\geq K$) symbols is allocated to UL pilots for channel estimation in each coherence block, as shown in Figs. 1(a) and 2. In other words, let $\mathbf{p}_k \in \mathbb{C}^P$ denote the ChEst pilot sequence used by the k^{th} user in any of the L cells. Then, for another user m in the same cell, we have $\mathbf{p}_k^H \mathbf{p}_m = P \delta_{km}$. We assume that the same P pilots are reused in each cell and each user is randomly allocated one of these pilots for channel estimation.

The pilot transmissions in all cells are assumed to be synchronized. Then, the received signal at BS j during pilot transmissions in the n^{th} coherence block (denoted as $\mathbf{Y}_j^{(p)}[n]$) can be written as

$$\mathbf{Y}_j^{(p)}[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\mu} \mathbf{h}_{jlk} \mathbf{p}_k^T + \mathbf{N}_j^{(p)}[n] \quad (2)$$

where $\mathbf{N}_j^{(p)}[n] \in \mathbb{C}^{M \times P}$ is the noise during pilot transmission.

We consider LMMSE and element-wise LMMSE techniques for estimating \mathbf{h}_{jlk} from the observed signal $\mathbf{Y}_j^{(p)}$ given in (2). In what follows, we first review these estimation techniques.

1) *LMMSE Channel Estimation*: From (2), the least-squares (LS) channel estimate of user (j, u) at BS j in the n^{th} coherent block (denoted as $\hat{\mathbf{h}}_{jju}^{LS}[n]$) can be obtained as follows³

$$\begin{aligned} \hat{\mathbf{h}}_{jju}^{LS}[n] &= \arg \min_{\mathbf{g}} \|\mathbf{Y}_j^{(p)}[n] - \sqrt{\mu} \mathbf{g} \mathbf{P}_u^{\mathbf{I}}\|^2 \\ &= \mathbf{h}_{jju} + \sum_{l \neq j} \mathbf{h}_{jlu} + \frac{1}{P\sqrt{\mu}} \mathbf{N}_j^{(p)}[n] \mathbf{p}_u^*. \end{aligned} \quad (3)$$

Using this LS channel estimate that serves as a sufficient statistic for \mathbf{h}_{jju} , the resultant LMMSE estimate can be easily derived to be [15]

$$\begin{aligned} \hat{\mathbf{h}}_{jju}^{LMMSE}[n] &= \mathbf{R}_{jju} \mathbf{Q}_{ju}^{-1} \hat{\mathbf{h}}_{jju}^{LS}[n] \\ \mathbf{Q}_{ju} &\triangleq \mathbb{E}\{\hat{\mathbf{h}}_{jju}^{LS}[n] (\hat{\mathbf{h}}_{jju}^{LS}[n])^H\} = \sum_{l=1}^L \mathbf{R}_{jlu} + \frac{1}{P\mu} \mathbf{I}. \end{aligned} \quad (4)$$

Although the channel estimates in the above equation assume that the covariance information is known, in practice it has to be estimated at the BS. Therefore, it is reasonable to replace these matrices with estimated covariance matrices ($\hat{\mathbf{R}}_{jju}$, and $\hat{\mathbf{Q}}_{ju}$) to get LMMSE-type channel estimate as follows

$$\hat{\mathbf{h}}_{jju}[n] = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1} \hat{\mathbf{h}}_{jju}^{LS}[n] \quad (5)$$

For known covariance case, the computational complexity in evaluating (4) is $\mathcal{O}(M^3)$. Furthermore, the computational complexity of a sample covariance matrix of an $M \times 1$ channel vector is $\mathcal{O}(NM^2)$, where N is the number of samples. Therefore, the total computational complexity involved in evaluating (5) is $\mathcal{O}(M^3 + M^2 N_R + M^2 N_Q)$, where N_R and N_Q are the number of pilot sequences (samples) used for computing $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$, respectively.

2) *Element-Wise LMMSE Channel Estimation*: An alternative approach for LMMSE channel estimation is to use the element-wise LMMSE estimate; this technique requires a fewer number of samples/pilots for the covariance estimation that does not grow with M [10].

The element-wise LMMSE estimate of the channel can be obtained as

$$[\hat{\mathbf{h}}_{jju}^{\text{el-LMMSE}}[n]]_p = \frac{[\mathbf{S}_{jju}]_{pp}}{[\mathbf{P}_{ju}]_{pp}} [\hat{\mathbf{h}}_{jju}^{LS}[n]]_p, \quad p \in \{1, \dots, M\}$$

where $\mathbf{S}_{jju} \triangleq \text{diag}(\mathbf{R}_{jju})$ and $\mathbf{P}_{ju} \triangleq \text{diag}(\mathbf{Q}_{ju})$. The element-wise LMMSE-type estimate with estimated covariance matrices ($\hat{\mathbf{S}}_{jju}$, and $\hat{\mathbf{P}}_{ju}$) can be written as

$$[\hat{\mathbf{h}}_{jju}^{\text{el}}[n]]_p = \frac{[\hat{\mathbf{S}}_{jju}]_{pp}}{[\hat{\mathbf{P}}_{ju}]_{pp}} [\hat{\mathbf{h}}_{jju}^{LS}[n]]_p, \quad p \in \{1, \dots, M\} \quad (6)$$

³As the channel observations, in this case, are linear measurements in gaussian noise, one should note that this is also an MMSE estimator.

Here, each diagonal element of $\hat{\mathbf{S}}_{jju}$ ($\hat{\mathbf{P}}_{ju}$) is computed using a sample variance estimator of the corresponding element of the channel vector. If we use N_R (N_Q) number of channel samples for estimating $\hat{\mathbf{S}}_{jju}$ ($\hat{\mathbf{P}}_{ju}$), the computational complexity involved in evaluating each element of $\hat{\mathbf{h}}_{jju}^{\text{el}}[n]$ is $\mathcal{O}(N_R + N_Q)$. Therefore, the total computational complexity involved in evaluating $\hat{\mathbf{h}}_{jju}^{\text{el}}$ is $\mathcal{O}(MN_R + MN_Q)$. Element-wise LMMSE-type channel estimation substantially reduces the computational complexity at the cost of some performance degradation caused due to the fact that we ignore non-diagonal elements of $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$. Later in Section V, we compare the performance of these two channel estimation methods using simulations.

B. Covariance Matrix Estimation

Several methods to address the covariance matrix estimation problem have been proposed in literature [11], [14]–[16]. However, among these methods, only the estimators in [15] and [16] are in closed-form and consequently, lend themselves to mathematical analysis. Moreover, since [16] is seen to outperform [15], we select the estimator in [16] for performance analysis when the estimate is used for both LMMSE-type and element-wise LMMSE-type channel estimation. We assume that the BSs have knowledge of the random phase sequences.

In this subsection, we briefly review the pilot structure introduced in [16] and the corresponding spatial covariance estimation method. The objective is to compute a pair of $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$ (or $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$) for each set of τ_s contiguous coherence blocks.

To obtain $\hat{\mathbf{Q}}_{ju}$, since the matrix \mathbf{Q}_{ju} is defined as the covariance matrix of $\hat{\mathbf{h}}_{jju}^{LS}[n]$, we use these LS channel estimates from multiple coherence blocks in a sample covariance estimator. We use a set of N_Q ($\geq M$) number of LS estimates for computing $\hat{\mathbf{Q}}_{ju}$. Therefore, we have the following unbiased covariance estimator $\hat{\mathbf{Q}}_{ju} = \frac{1}{N_Q} \sum_{n=1}^{N_Q} \hat{\mathbf{h}}_{jju}^{LS}[n] (\hat{\mathbf{h}}_{jju}^{LS}[n])^H$. Similarly, the unbiased estimate of \mathbf{P}_{ju} is obtained using a sample covariance estimator as follows

$$[\hat{\mathbf{P}}_{ju}]_{pp} = \frac{1}{N_Q} \sum_{n=1}^{N_Q} |[\hat{\mathbf{h}}_{jju}^{LS}[n]]_p|^2, \quad \forall p \in 1 \dots M.$$

For estimating \mathbf{R}_{jju} and \mathbf{S}_{jju} , as depicted by the red coherence blocks in Fig. 1(b), each user transmits an additional pilot sequence of length P symbols for N_R out of the τ_s coherence blocks. Specifically, the CovEst pilots, denoted as $\{\phi_{lk}[n]\}_{n=1}^{N_R}$, are transmitted by the user (l, k) , with the pilot sequence in n^{th} coherence block given as a phase-shifted version of the ChEst pilot, i.e., $\phi_{lk}[n] = e^{j\theta_{ln}} \mathbf{p}_k$. The phase-shifts $\{\theta_{ln}\}_{n=1}^{N_R}$ are (pseudo)-random and are generated such that $\{\theta_{ln}\}_{n=1}^{N_R}$ is independent of the channel vectors and satisfies $\mathbb{E}\{e^{j\theta_{ln}}\} = 0$ [16]. A random sequence that satisfies both these properties is $\theta_{ln} \sim \mathcal{U}[0, 2\pi)$. Furthermore, the random phase sequences are assumed to be i.i.d across cells.

Now, let $\mathbf{Y}_j^{(r)}[n]$ be the received signal when the users transmit the CovEst pilots $\phi_{ju}[n]$. Then, $\mathbf{Y}_j^{(r)}[n]$ can be

written as

$$\mathbf{Y}_j^{(r)}[n] = \sum_{l=1}^L \sum_{k=1}^K \sqrt{\mu} \mathbf{h}_{jlk} \phi_{lk}^T[n] + \mathbf{N}_j^{(r)}[n] \quad (7)$$

where $\mathbf{N}_j^{(r)}[n]$ is the AWGN noise at the BS that has the same statistics as $\mathbf{N}_j^{(p)}[n]$.

We denote LS channel estimates obtained from the pilots \mathbf{p}_u and ϕ_{ju} as $\hat{\mathbf{h}}_{jju}^{(1)}[n]$ and $\hat{\mathbf{h}}_{jju}^{(2)}[n]$, respectively. Using (2) and (7) and by using the fact that $\phi_{lk}[n] = e^{j\theta_{ln}} \mathbf{p}_k$, the LS estimates can be straightforwardly obtained as [16]

$$\hat{\mathbf{h}}_{jju}^{(1)}[n] = \mathbf{h}_{jju} + \sum_{l \neq j} \mathbf{h}_{jlu} + \frac{1}{P\sqrt{\mu}} \mathbf{N}_j^{(p)}[n] \mathbf{p}_u^* \quad (8)$$

$$\begin{aligned} \hat{\mathbf{h}}_{jju}^{(2)}[n] &= \mathbf{h}_{jju} + \sum_{l \neq j} \mathbf{h}_{jlu} e^{j(\theta_{ln} - \theta_{jn})} \\ &\quad + \frac{1}{P\sqrt{\mu}} \mathbf{N}_j^{(r)}[n] \mathbf{p}_u^* e^{-j\theta_{jn}} \end{aligned} \quad (9)$$

In the following subsections, we describe both cases of complete and diagonal matrix estimation using the aforementioned LS channel estimates.

1) *Estimation of $\hat{\mathbf{R}}_{jju}$.* Note that the second and third terms in (8), corresponding to the interference and noise, respectively, are independent of the second and third terms in (9). Consequently, the cross-correlation of $\hat{\mathbf{h}}_{jju}^{(1)}[n]$ and $\hat{\mathbf{h}}_{jju}^{(2)}[n]$ can be easily shown to be same as the covariance matrix \mathbf{R}_{jju} [16]. Therefore, we can use the following unbiased Hermitian-symmetric sample cross-covariance matrix as an estimate for \mathbf{R}_{jju}

$$\begin{aligned} \hat{\mathbf{R}}_{jju} &= \frac{1}{2N_R} \sum_{n=1}^{N_R} \hat{\mathbf{h}}_{jju}^{(1)}[n] \left(\hat{\mathbf{h}}_{jju}^{(2)}[n] \right)^H \\ &\quad + \frac{1}{2N_R} \sum_{n=1}^{N_R} \hat{\mathbf{h}}_{jju}^{(2)}[n] \left(\hat{\mathbf{h}}_{jju}^{(1)}[n] \right)^H \end{aligned} \quad (10)$$

As $N_R \rightarrow \infty$, one can show that the estimated covariance matrix converges in probability to the true covariance matrix, i.e., $\hat{\mathbf{R}}_{jju} \xrightarrow[N_R \rightarrow \infty]{P} \mathbf{R}_{jju}$. However, this unbiased covariance estimator does not guarantee positive diagonal elements for finite N_R . Therefore, we consider a regularized estimate for the covariance matrix given by

$$\hat{\mathbf{R}}_{jju} = \alpha_R \hat{\mathbf{R}}_{jju} + (1 - \alpha_R) \mathbf{R}_b \quad (11)$$

where \mathbf{R}_b is an arbitrary symmetric positive definite bias-matrix, and α_R is a design parameter. Additionally, it is useful to define $\hat{\mathbf{R}}_{jju}$ to denote the expected value of $\hat{\mathbf{R}}_{jju}$ as $\hat{\mathbf{R}}_{jju} \triangleq \mathbb{E}\{\hat{\mathbf{R}}_{jju}\} = \alpha_R \mathbf{R}_{jju} + (1 - \alpha_R) \mathbf{R}_b$.

2) *Estimation of $\hat{\mathbf{S}}_{jju}$.* For element-wise LMMSE-type estimation, it is sufficient to estimate the diagonal matrix \mathbf{S}_{jju} . Therefore, we use an unbiased Hermitian-symmetric covariance estimate $\hat{\mathbf{S}}_{jju}$ (similar to $\hat{\mathbf{R}}_{jju}$) as follows

$$\begin{aligned} &[\hat{\mathbf{S}}_{jju}]_{pp} \\ &= \frac{1}{2N_R} \sum_{n=1}^{N_R} [\hat{\mathbf{h}}_{jju}^{(1)}[n]]_p [\hat{\mathbf{h}}_{jju}^{(2)}[n]]_p^* \\ &\quad + \frac{1}{2N_R} \sum_{n=1}^{N_R} [\hat{\mathbf{h}}_{jju}^{(2)}[n]]_p [\hat{\mathbf{h}}_{jju}^{(1)}[n]]_p^*, \quad \forall p \in 1 \dots M. \end{aligned} \quad (12)$$

A regularized estimate for \mathbf{S}_{jju} is given by

$$\hat{\mathbf{S}}_{jju} = \alpha_R \hat{\mathbf{S}}_{jju} + (1 - \alpha_R) \text{diag}(\mathbf{R}_b). \quad (13)$$

We define $\bar{\mathbf{S}}_{jju}$ as the expected value of $\hat{\mathbf{S}}_{jju}$, $\bar{\mathbf{S}}_{jju} \triangleq \mathbb{E}\{\hat{\mathbf{S}}_{jju}\} = \alpha_R \mathbf{S}_{jju} + (1 - \alpha_R) \text{diag}(\mathbf{R}_b)$, for future use.

In summary, the BS needs to compute channel covariance matrices for each set of τ_s coherence blocks in order to obtain the LMMSE-type/element-wise LMMSE-type channel estimates in each coherence block within the set.

III. AVERAGE UL AND DL SPECTRAL EFFICIENCIES AS PERFORMANCE METRIC

In order to analyze the performance of the covariance estimation algorithm in a massive MIMO system, we derive a closed-form expression for a performance metric that captures the impact of pilot overhead. Clearly, the performance of a massive MIMO system directly depends on the quality of channel covariance estimates. In the literature [14]–[16], the achievable SE value is typically computed for a single set of τ_s coherence blocks as a function of the estimated covariance matrices corresponding to that set. Such an SE value corresponding to a single realization of covariance matrices is particularly important for designing practical receivers but does not clearly capture the impact of covariance estimation. Therefore, we consider average SE computed across different realizations of the covariance matrices (ensemble average) as the performance metric. We derive closed-form expressions for average SE for both UL and DL data for LMMSE-type and element-wise LMMSE-type channel estimation.⁴ Note that the maximum ratio combining corresponds to a lower SE value when compared to LMMSE combining [10]. Moreover, the aim of the average SE-based performance metric derived in this paper is to capture the impact pilot overhead but not to present an achievable rate. Therefore, we use matched filter precoding and maximum ratio combining for deriving the SE expressions in the DL and UL communication, respectively.

A. Uplink Spectral Efficiency

In this section, the average SE for the UL channel of a target user (j, u) is derived when the channel estimates are used in a maximum ratio combiner at the BS. For the maximum ratio combiner, the combining vector $\mathbf{v}_{ju}[n]$ can be written as $\mathbf{v}_{ju}[n] = \hat{\mathbf{h}}_{jju}[n] = \hat{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{jju}^{LS}[n]$, where

$$\hat{\mathbf{W}}_{ju} = \begin{cases} \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}, & \text{LMMSE-type channel estimate} \\ \hat{\mathbf{S}}_{jju} \hat{\mathbf{P}}_{ju}^{-1}, & \text{element-wise LMMSE-type} \\ & \text{channel estimate.} \end{cases}$$

For the sake of mathematical tractability, we make the following assumptions

- $\hat{\mathbf{R}}_{jju}$ ($\hat{\mathbf{S}}_{jju}$) and $\hat{\mathbf{Q}}_{ju}$ ($\hat{\mathbf{P}}_{ju}$) are each computed from a different non-overlapping set of coherence blocks that does not include n^{th} block [15]. Consequently, the random variables $\hat{\mathbf{R}}_{jju}/\hat{\mathbf{S}}_{jju}$, $\hat{\mathbf{Q}}_{ju}/\hat{\mathbf{P}}_{ju}$, and $\hat{\mathbf{h}}_{jju}^{LS}[n]$ are mutually uncorrelated.

⁴Note that, [14]–[16] utilize numerically computed average rate as a performance measure for covariance matrix estimation method.

- For the LMMSE-type channel estimate, N_Q is assumed greater than M , so that the distribution of $\hat{\mathbf{Q}}_{ju}^{-1}$ is non-degenerate inverse Wishart.

The received combined signal is given by

$$\begin{aligned} \mathbf{v}_{ju}^H \mathbf{y}_j &= \sqrt{\mu} \mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\} x_{ju} \\ &+ \sqrt{\mu} (\mathbf{v}_{ju}^H \mathbf{h}_{jju} - \mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\}) x_{ju} \\ &+ \sum_{k \neq u} \sqrt{\mu} \mathbf{v}_{ju}^H \mathbf{h}_{jjk} x_{jk} \\ &+ \sum_{l \neq j} \sum_{k=1}^K \sqrt{\mu} \mathbf{v}_{ju}^H \mathbf{h}_{jlk} x_{lk} + \mathbf{v}_{ju}^H \mathbf{n}_j \end{aligned} \quad (14)$$

In (14), the first term corresponds to the signal component, the second term is a result of the uncertainty in the array gain, the third term corresponds to the non-coherent intracell interference, the fourth term corresponds to the coherent interference from pilot contamination, and the last term corresponds to the additive noise component. Since the first term is uncorrelated with the subsequent terms, a lower bound on SE of the UL channel from user (j, u) to BS j can be obtained as [15]

$$SE_{ju}^{(ul)} = \left(1 - \frac{P}{C_u} - \frac{N_R P}{C_u \tau_s}\right) \log_2 \left(1 + \gamma_{ju}^{(ul)}\right), \quad [\text{bits/s/Hz}]$$

where $\gamma_{ju}^{(ul)}$ is given by

$$\gamma_{ju}^{(ul)} = \frac{|\mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\}|^2}{\sum_{l,k} \mathbb{E}\{|\mathbf{v}_{ju}^H \mathbf{h}_{jlk}|^2\} - |\mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{h}_{jju}\}|^2 + \frac{1}{\mu} \mathbb{E}\{\mathbf{v}_{ju}^H \mathbf{v}_{ju}\}}$$

and the expectation $\mathbb{E}\{\cdot\}$ is over the channel realizations. In the pre-log factor, P/C_u accounts for ChEst pilots, and $N_R P/C_u \tau_s$ accounts for CovEst pilots. However, since we assume that $\hat{\mathbf{W}}_{ju}$ and $\hat{\mathbf{h}}_{jju}^{LS}[n]$ are mutually independent, we have $\mathbb{E}\{\cdot\} = \mathbb{E}_W \{\mathbb{E}_{h,LS}\{\cdot\}\}$, where \mathbb{E}_W is the expectation over $\hat{\mathbf{W}}_{ju}$, and $\mathbb{E}_{h,LS}$ is the expectation over the LS estimate.

Let $\mathbf{R}_s \triangleq \sum_{l=1}^L \sum_{k=1}^K \mathbf{R}_{jlk} + \frac{1}{\mu} \mathbf{I}$. Then, the signal to interference plus noise ratio (SINR) expression can be further simplified to (15) [15], as shown at the bottom of the next page.

B. Uplink Spectral Efficiency When $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}$

In this subsection, expressions for all the terms given in (15) are derived for the case when $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}$. In what follows, $\mathbb{E}_R\{\cdot\}$ represents the expectation over $\hat{\mathbf{R}}_{jju}$, $\mathbb{E}_Q\{\cdot\}$ represents the expectation over $\hat{\mathbf{Q}}_{ju}$, and $\mathbb{E}_W\{\cdot\}$ represents the expectation over both $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$. It should be noted that, as already mentioned, we have assumed that $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$ are estimated from different pilot resources (coherence blocks) such that the estimates are independent to each other. Therefore, $\mathbb{E}_R\{\cdot\}$ and $\mathbb{E}_Q\{\cdot\}$ can be evaluated independently.

Before analytically deriving the expectations for the terms in (15), we present some useful lemmas.

Lemma 1: Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$, and for any mutually independent M -dimensional random vector \mathbf{h} distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R})$, we have

$$\mathbb{E}\{\mathbf{h} \mathbf{h}^H \mathbf{A} \mathbf{h} \mathbf{h}^H\} = \mathbf{R} \mathbf{A} \mathbf{R} + \mathbf{R} \text{tr}(\mathbf{A} \mathbf{R}) \quad (17)$$

$$\mathbb{E}\{|\mathbf{h}^H \mathbf{A} \mathbf{h}|^2\} = |\text{tr}(\mathbf{A}^H \mathbf{R})|^2 + \text{tr}(\mathbf{A} \mathbf{R} \mathbf{A}^H \mathbf{R}). \quad (18)$$

Proof: The proof is available in Appendix A. ■

Lemma 2: Given a Hermitian matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$, an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$, and a complex Wishart matrix, $\mathbf{X} \in \mathbb{C}^{M \times M}$, distributed as $\mathcal{W}(N, \mathbf{I})$, we have

$$\mathbb{E}\{[\mathbf{X}^{-1}]_{ij}\} = \frac{[\mathbf{I}]_{ij}}{N-M} \quad (19)$$

$$\mathbb{E}\{[\mathbf{X}^{-1}]_{ij} [\mathbf{X}^{-1}]_{lk}\} = \frac{[\mathbf{I}]_{ij} [\mathbf{I}]_{lk} + \frac{1}{N-M} [\mathbf{I}]_{lj} [\mathbf{I}]_{ik}}{(N-M)^2 - 1} \quad (20)$$

$$\mathbb{E}\{\text{tr}(\mathbf{X}^{-2} \mathbf{C})\} = \frac{N}{(N-M)^3 - (N-M)} \text{tr}(\mathbf{C}) \quad (21)$$

$$\mathbb{E}\{|\text{tr}(\mathbf{X}^{-1} \mathbf{A})|^2\} = \frac{|\text{tr}(\mathbf{A})|^2 + \frac{1}{N-M} \text{tr}(\mathbf{A} \mathbf{A}^H)}{(N-M)^2 - 1}. \quad (22)$$

Proof: The proof is available in Appendix B. ■

Lemma 3: Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$, we have

$$\begin{aligned} &\mathbb{E}\{\ddot{\mathbf{R}}_{jju} \mathbf{A} \ddot{\mathbf{R}}_{jju}\} \\ &= \mathbf{R}_{jju} \mathbf{A} \mathbf{R}_{jju} \\ &+ \frac{1}{2N_R} \mathbf{Q}_{ju} \text{tr}(\mathbf{A} \mathbf{Q}_{ju}) + \frac{1}{2N_R} \mathbf{R}_{jju} \text{tr}(\mathbf{A} \mathbf{R}_{jju}) \end{aligned} \quad (23)$$

$$\begin{aligned} &\mathbb{E}\{|\text{tr}(\ddot{\mathbf{R}}_{jju} \mathbf{A})|^2\} \\ &= |\text{tr}(\mathbf{R}_{jju} \mathbf{A})|^2 \\ &+ \frac{1}{2N_R} \left\{ \text{tr}(\mathbf{A} \mathbf{Q}_{ju} \mathbf{A}^H \mathbf{Q}_{ju}) + \text{tr}(\mathbf{A} \mathbf{R}_{jju} \mathbf{A}^H \mathbf{R}_{jju}) \right\} \end{aligned} \quad (24)$$

Proof: The proof of this lemma uses Lemma 1 and is presented in Appendix C. ■

Now we are ready to formulate the key theorem of this subsection.

Theorem 1: The numerator term of (15) when $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}$ is given by

$$\begin{aligned} &\mathbb{E}_W \left\{ \text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) \right\} \\ &= \text{tr}(\mathbf{W}_{ju}^H \mathbf{R}_{jju}) \\ &+ \left\{ \frac{N_Q}{N_Q - M} \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) - \text{tr}(\mathbf{W}_{ju}^H \mathbf{R}_{jju}) \right\} \end{aligned} \quad (25)$$

The first and second terms of the denominator in (15) are given by

$$\begin{aligned} &\mathbb{E}_W \left\{ \text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s) \right\} \\ &= \text{tr}(\mathbf{W}_{ju} \mathbf{Q}_{ju} \mathbf{W}_{ju}^H \mathbf{R}_s) \\ &+ \left\{ \kappa_1 \text{tr}(\bar{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \bar{\mathbf{W}}_{ju}^H \mathbf{R}_s) - \text{tr}(\mathbf{W}_{ju} \mathbf{Q}_{ju} \mathbf{W}_{ju}^H \mathbf{R}_s) \right. \\ &+ \left. \frac{\alpha_R^2 \kappa_1}{2N_R} M \text{tr}(\mathbf{R}_s \mathbf{Q}_{ju}) + \frac{\alpha_R^2 \kappa_1}{2N_R} \text{tr}(\mathbf{W}_{ju}) \text{tr}(\mathbf{R}_s \mathbf{R}_{jju}) \right\} \end{aligned} \quad (26)$$

and (16), as shown at the bottom of the next page, respectively. Here, $\kappa_1 \triangleq N_Q \kappa_2 / (N_Q - M)$, $\kappa_2 \triangleq N_Q^2 / ((N_Q - M)^2 - 1)$, $\bar{\mathbf{W}}_{ju} \triangleq \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}$ and $\bar{\mathbf{W}}_{lu} \triangleq \hat{\mathbf{R}}_{jlu} \hat{\mathbf{Q}}_{ju}^{-1}$ for all $l = 1$ to L .

Proof: We define a matrix $\tilde{\mathbf{Q}}_{ju}$ as

$$\tilde{\mathbf{Q}}_{ju} \triangleq N_Q (\mathbf{Q}_{ju}^{-\frac{1}{2}} \hat{\mathbf{Q}}_{ju} \mathbf{Q}_{ju}^{-\frac{1}{2}}). \quad (27)$$

It can be seen that $\tilde{\mathbf{Q}}_{ju}$ is Wishart distributed, i.e., $\mathcal{W}(N_Q, \mathbf{I})$.

Using (27) and the fact that $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}$, the numerator term of (15) can be written as

$$\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\} = N_Q \mathbb{E}_W\{\text{tr}(\mathbf{Q}_{ju}^{-\frac{1}{2}} \tilde{\mathbf{Q}}_{ju}^{-1} \mathbf{Q}_{ju}^{-\frac{1}{2}} \hat{\mathbf{R}}_{jju} \mathbf{R}_{jju})\}. \quad (28)$$

Taking direct expectation over $\hat{\mathbf{R}}_{jju}$ in (28) and also using Lemma 2, (25) can be obtained.

Proof of (26) and (16) is as follows. Substituting $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}$ into the first and second terms in the denominator of (15) and using Lemma 2, we get the following equations

$$\begin{aligned} & \mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s)\} \\ &= \kappa_1 \mathbb{E}_R\{\text{tr}(\mathbf{Q}_{ju}^{-1} \hat{\mathbf{R}}_{jju} \mathbf{R}_s \hat{\mathbf{R}}_{jju})\} \\ & \mathbb{E}_W\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2\} \\ &= \kappa_2 \mathbb{E}_R\{|\text{tr}(\mathbf{Q}_{ju}^{-1} \hat{\mathbf{R}}_{jju} \mathbf{R}_{jlu})|^2\} \\ &+ \frac{\kappa_1}{N_Q} \mathbb{E}_R\{\text{tr}(\mathbf{Q}_{ju}^{-1} \hat{\mathbf{R}}_{jju} \mathbf{R}_{jlu}^2 \hat{\mathbf{R}}_{jju} \mathbf{Q}_{ju}^{-1})\}. \end{aligned} \quad (29) \quad (30)$$

Then using Lemma 3, and substituting (11) into (29) and (30), we get (26) and (16), respectively. ■

Note that the expectation terms given in Theorem 1 contain two components: (i) the component that corresponds to known covariance information (first term of the right-hand side of the equations) and (ii) a penalty component (all terms except the first term of the right-hand side of the equations) due to regularization of \mathbf{R}_{jju} estimate and due to imperfect channel covariance information. Note that for $\alpha_R = 1$, and as N_R and N_Q tend to infinity, the penalty components of the expectation terms vanish.

C. Uplink Spectral Efficiency When $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju} \hat{\mathbf{P}}_{ju}^{-1}$

In this subsection, we present the derivations for all the terms given in (15) when $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju} \hat{\mathbf{P}}_{ju}^{-1}$. In what follows, $\mathbb{E}_S\{\cdot\}$ represents the expectation over $\hat{\mathbf{S}}_{jju}$, $\mathbb{E}_P\{\cdot\}$ represents the expectation over $\hat{\mathbf{P}}_{ju}$, and $\mathbb{E}_W\{\cdot\}$ represents the expectation over both $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$.

Before analytically deriving the expectations for the terms in (15), we present some useful lemmas.

Lemma 4: Given a zero mean complex Gaussian 2×1 random vector $\mathbf{h} = [h_1, h_2]^T$ with covariance matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

we can state that $\mathbb{E}\{|h_1|^2|h_2|^2\} = r_{11}r_{22} + r_{12}r_{21}$.

Proof: The proof of this lemma is straight forward to obtain and we omit it due to lack of space. ■

Lemma 5: Given arbitrary matrices $\mathbf{A}_1 \in \mathbb{C}^{M \times M}$, $\mathbf{A}_2 \in \mathbb{C}^{M \times M}$, $\mathbf{A} \in \mathbb{C}^{M \times M}$, and a matrix $\mathbf{Y} = \mathbf{Z}/2$, where \mathbf{Z} is a diagonal matrix whose elements are i.i.d. χ^2 random variables with $2N$ -degrees of freedom ($N > 2$), we have

$$\mathbb{E}\{\text{tr}(\mathbf{Y}^{-1} \mathbf{A}_1 \mathbf{Y}^{-1} \mathbf{A}_2)\} = \tau_1 \text{tr}(\mathbf{A}_1 \mathbf{A}_2) + \tau_2 \text{tr}(\mathbf{A}_{1d} \mathbf{A}_{2d}) \quad (31)$$

$$\mathbb{E}\{|\text{tr}(\mathbf{Y}^{-1} \mathbf{A})|^2\} = \tau_1 |\text{tr}(\mathbf{A})|^2 + \tau_2 \text{tr}(\mathbf{A}_d^H \mathbf{A}_d) \quad (32)$$

where $\tau_1 \triangleq 1/(N-1)^2$, $\tau_2 \triangleq \tau_1/(N-2)$, $\mathbf{A}_{1d} \triangleq \text{diag}(\mathbf{A}_1)$, $\mathbf{A}_{2d} \triangleq \text{diag}(\mathbf{A}_2)$, and $\mathbf{A}_d \triangleq \text{diag}(\mathbf{A})$.

Proof: The proof is available in Appendix D. ■

Lemma 6: Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$ and an arbitrary diagonal matrix $\mathbf{D} \in \mathbb{R}^{M \times M}$, then

$$\begin{aligned} \mathbb{E}\{\ddot{\mathbf{S}}_{jju} \mathbf{A} \ddot{\mathbf{S}}_{jju}\} &= \mathbf{S}_{jju} \mathbf{A} \mathbf{S}_{jju} + \frac{1}{2N_R} \mathbf{A} \circ \mathbf{Q}_{ju} \circ \mathbf{Q}_{ju} \\ &+ \frac{1}{2N_R} \mathbf{A} \circ \mathbf{R}_{jju} \circ \mathbf{R}_{jju} \end{aligned} \quad (33)$$

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\ddot{\mathbf{S}}_{jju} \mathbf{D})|^2\} &= |\text{tr}(\mathbf{S}_{jju} \mathbf{D})|^2 \\ &+ \frac{1}{2N_R} \sum_{p=1}^M \sum_{q=1}^M [\mathbf{D}(\mathbf{Q}_{ju} \circ \mathbf{Q}_{ju})\mathbf{D}]_{pq} \\ &+ \frac{1}{2N_R} \sum_{p=1}^M \sum_{q=1}^M [\mathbf{D}(\mathbf{R}_{jju} \circ \mathbf{R}_{jju})\mathbf{D}]_{pq} \end{aligned} \quad (34)$$

Proof: The proof is available in Appendix E. ■

Now we are ready to formulate the key theorem of this subsection.

Theorem 2: The numerator term of (15) when $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju} \hat{\mathbf{P}}_{ju}^{-1}$ is given by

$$\begin{aligned} & \mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\} \\ &= \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) \\ &+ \left\{ \frac{N_Q}{(N_Q-1)} \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) - \text{tr}(\mathbf{W}_{ju}^H \mathbf{R}_{jju}) \right\} \end{aligned} \quad (35)$$

$$\gamma_{ju}^{(\text{ul})} = \frac{|\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\}|^2}{\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s)\} + \sum_{l=1}^L \mathbb{E}_W\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2\} - |\mathbb{E}_W\{\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\}|^2} \quad (15)$$

$$\begin{aligned} \mathbb{E}_W\{|\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2\} &= |\text{tr}(\mathbf{W}_{ju}^H \mathbf{R}_{jlu})|^2 + \left\{ \kappa_2 |\text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2 - |\text{tr}(\mathbf{W}_{ju}^H \mathbf{R}_{jlu})|^2 + \frac{\alpha_R^2 \kappa_2}{2N_R} \text{tr}(\mathbf{W}_{lu} \mathbf{Q}_{ju} \mathbf{W}_{lu}^H \mathbf{Q}_{ju}) \right. \\ &+ \frac{\alpha_R^2 \kappa_2}{2N_R} \text{tr}(\mathbf{W}_{lu} \mathbf{R}_{jju} \mathbf{W}_{lu}^H \mathbf{R}_{jju}) + \frac{\kappa_1}{N_Q} \text{tr}(\bar{\mathbf{W}}_{ju}^H \bar{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \mathbf{W}_{lu}^H \mathbf{W}_{lu} \mathbf{Q}_{ju}) + \frac{\alpha_R^2 \kappa_1}{2N_Q N_R} M \text{tr}(\mathbf{W}_{jlu}^2 \mathbf{Q}_{ju}^2) \\ &\left. + \frac{\alpha_R^2 \kappa_1}{2N_Q N_R} \text{tr}(\mathbf{W}_{ju}) \text{tr}(\mathbf{W}_{jlu}^2 \mathbf{Q}_{ju} \mathbf{R}_{jju}) \right\} \end{aligned} \quad (16)$$

The first and second terms of the denominator in (15) are given by (36) and (37), as shown at the bottom of the next page, respectively, where $\kappa_3 = N_Q^2/(N_Q - 1)^2$, $\kappa_4 = \kappa_3/(N_Q - 2)$, $\mathbf{S}_s \triangleq \text{diag}(\mathbf{R}_s)$, $\tilde{\mathbf{W}}_{ju} \triangleq \hat{\mathbf{S}}_{jju} \mathbf{P}_{ju}^{-1}$ and $\tilde{\mathbf{W}}_{lu} \triangleq \hat{\mathbf{S}}_{jlu} \mathbf{P}_{ju}^{-1}$ for all $l = 1$ to L .

Proof: We define the diagonal matrix $\tilde{\mathbf{P}}_{ju}$ as follows

$$\tilde{\mathbf{P}}_{ju} \triangleq N_Q (\mathbf{P}_{ju}^{-1} \hat{\mathbf{P}}_{ju}). \quad (39)$$

It can be seen that the elements of $2\tilde{\mathbf{P}}_{ju}$ are i.i.d. χ^2 random variables with $2N$ -degrees of freedom. Using (39) and the fact that $\tilde{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju} \tilde{\mathbf{P}}_{ju}^{-1}$, the numerator term of (15) can be written as

$$\begin{aligned} & \mathbb{E}_W \{ \text{tr}(\tilde{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) \} \\ &= N_Q \mathbb{E}_W \{ \text{tr}(\tilde{\mathbf{P}}_{ju}^{-1} \mathbf{P}_{ju}^{-1} \hat{\mathbf{S}}_{jju} \mathbf{R}_{jju}) \} \\ &= N_Q \sum_{p=1}^M \mathbb{E}_P \{ [\tilde{\mathbf{P}}_{ju}^{-1}]_{pp} \} \mathbb{E}_S \{ [\mathbf{P}_{ju}^{-1} \hat{\mathbf{S}}_{jju} \mathbf{R}_{jju}]_{pp} \}. \end{aligned} \quad (40)$$

Taking direct expectation over $\hat{\mathbf{S}}_{jju}$ in (40) and using the properties of inverse χ^2 distribution, (35) can be obtained.

Proof of (36) and (37) is as follows. Substituting $\tilde{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju} \tilde{\mathbf{P}}_{ju}^{-1}$ and (39) into the first and second denominator terms of (15) and using Lemma 5, we get the following equations

$$\begin{aligned} & \mathbb{E}_W \{ \text{tr}(\tilde{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \tilde{\mathbf{W}}_{ju}^H \mathbf{R}_s) \} \\ &= \kappa_3 \mathbb{E}_S \{ \text{tr}(\mathbf{P}_{ju}^{-1} \mathbf{Q}_{ju} \mathbf{P}_{ju}^{-1} \hat{\mathbf{S}}_{jju} \mathbf{R}_s \hat{\mathbf{S}}_{jju}) \} \\ & \quad + \kappa_4 \mathbb{E}_S \{ \text{tr}(\mathbf{P}_{ju}^{-1} \hat{\mathbf{S}}_{jju} \mathbf{S}_s \hat{\mathbf{S}}_{jju}) \} \end{aligned} \quad (41)$$

$$\begin{aligned} & \mathbb{E}_W \{ |\text{tr}(\tilde{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2 \} \\ &= \kappa_3 \mathbb{E}_S \{ |\text{tr}(\mathbf{P}_{ju}^{-1} \hat{\mathbf{S}}_{jju} \mathbf{S}_{jlu})|^2 \} \\ & \quad + \kappa_4 \mathbb{E}_S \{ \text{tr}(\mathbf{P}_{ju}^{-2} \hat{\mathbf{S}}_{jju}^2 \mathbf{S}_{jlu}^2) \}. \end{aligned} \quad (42)$$

Then using Lemma 6 and substituting (13) into (41) and (42), we get (36) and (37), respectively. ■

Similar to Theorem 1, the penalty components of the expectation terms given in Theorem 2 also vanish if $\alpha_R = 1$, and as N_R and N_Q tend to infinity.

D. Downlink Spectral Efficiency

The DL spectral efficiency for user (j, u) is given in this section for a matched filter precoder, i.e., $\mathbf{b}_{ju} = \hat{\mathbf{h}}_{jju}[n] / \sqrt{\mathbb{E}\{\|\hat{\mathbf{h}}_{jju}[n]\|^2\}} = \tilde{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{jju}^{LS} / \sqrt{\mathbb{E}\{\|\tilde{\mathbf{W}}_{ju} \hat{\mathbf{h}}_{jju}^{LS}\|^2\}}$. Therefore, the received signal at user (j, u) can be written as

$$\begin{aligned} z_{ju} &= \sqrt{\lambda} \mathbb{E}\{\mathbf{b}_{ju}^H \mathbf{h}_{jju}\} d_{ju} \\ & \quad + \sqrt{\lambda} (\mathbf{b}_{ju}^H \mathbf{h}_{jju} - \mathbb{E}\{\mathbf{b}_{ju}^H \mathbf{h}_{jju}\}) d_{ju} \\ & \quad + \sum_{k \neq u} \sqrt{\lambda} (\mathbf{b}_{ju}^H \mathbf{h}_{jjk}) d_{jk} \\ & \quad + \sum_{l \neq j} \sum_{k=1}^K \sqrt{\lambda} (\mathbf{b}_{ju}^H \mathbf{h}_{ljk}) d_{lk} + e_{ju}. \end{aligned} \quad (43)$$

Here, we assume that the scalar in the denominator of the precoding vector, $\sqrt{\mathbb{E}\{\|\hat{\mathbf{h}}_{jju}[n]\|^2\}}$, is a known constant at the

BS. The first term in (43) corresponds to the desired signal component, the second term corresponds to the uncertainty in the DL transmit array gain, the third term corresponds to the non-coherent intra-cell interference, the coherent interference from pilot contamination given by the fourth term, and the last term represents the additive noise component. The second term in (43) corresponds to the uncertainty in the DL transmit array gain. Then, due to the similarity between the UL received combined signal in (14) to the DL received signal, a lower bound on DL channel SE of the user (j, u) can be easily shown to be

$$\text{SE}_{ju}^{(dl)} = \log_2 \left(1 + \gamma_{ju}^{(dl)} \right) \quad [\text{bits/s/Hz}],$$

where $\gamma_{ju}^{(dl)}$ is given by (38), as shown at the bottom of the next page, and $\mathbf{R}_s^{(dl)} \triangleq \sum_{l=1}^L \sum_{k=1}^K \mathbf{R}_{jlk}$. We utilize channel reciprocity and avoid the use of DL pilots. Consequently, there is no pre-log factor for the SE expression. The expectation taken in all the terms of (38) is over the random matrix $\tilde{\mathbf{W}}_{ju}$. However, $\tilde{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju} \hat{\mathbf{Q}}_{ju}^{-1}$ for the LMMSE-type channel estimation and $\tilde{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju} \tilde{\mathbf{P}}_{ju}^{-1}$ for the element-wise LMMSE-type channel estimation. These expectation terms are already presented in Theorems 1, and 2 for the LMMSE-type, and element-wise LMMSE-type, respectively. Furthermore, \mathbf{R}_s should be replaced by $\mathbf{R}_s^{(dl)}$ in computing the expectation terms.

IV. MAIN DISCUSSION: IMPACT OF PILOT OVERHEAD AND CHANNEL ESTIMATION

Based on the closed-form SE expressions derived in the previous section, we have established a direct relation between the average SE value and the parameters N_R and N_Q . Thus, we discuss here the impact of these parameters on the SE corresponding to the LMMSE-type and element-wise LMMSE-type channel estimation. We also address the question of how to choose between LMMSE-type and element-wise LMMSE-type channel estimations.

It can be noted from the expectation terms in Theorems 1 and 2 that the penalty components due to imperfect covariance information gradually vanish with an increase in N_R and N_Q , but the penalty due to the regularization remains finite. Furthermore, if $\|\mathbf{W}_{ju} - \tilde{\mathbf{W}}_{ju}\| / \|\mathbf{W}_{ju}\| \ll 1$ (i.e., if α_R is close to 1), one can state that these expectation terms converge to the values that correspond to the known covariance case. However, despite leading to an improvement in $\gamma_{ju}^{(ul)}$ (due to convergence of the expected values), an increase in N_R also causes a degradation in the pre-log factor of the derived UL SE expression. Therefore, the choice of N_R impacts UL SE in two ways: (i) smaller the value of N_R , higher the error in covariance estimation and hence lower the value of UL SE and (ii) larger the value of N_R , higher the consumption of UL resources and hence lower the value of UL SE. Whereas, due to the absence of DL pilots, the DL SE does not degrade with an increase in N_R ; it gradually rises to the DL SE value that corresponds to the known covariance case. Larger N_Q makes both the UL and DL SE better due to the improved estimates of \mathbf{Q}_{ju} (or \mathbf{P}_{ju}). Therefore, given an SE requirement, the aim

here is to choose minimum N_R and N_Q values that are sufficient to provide the desired SE.

Since estimating \mathbf{Q}_{ju} (or \mathbf{P}_{ju}) does not involve additional pilot transmission, choosing N_Q is not as critical as choosing N_R . Therefore, if we consider N_Q as known, it is also important to derive N_R values that make the LMMSE-type channel estimation preferable to the element-wise LMMSE-type one, and vice-versa. By comparing the UL/DL SINR values (in (15) or (38)) for the two channel estimation techniques, we can compute a threshold, \bar{N}_R , such that the element-wise LMMSE-type estimator is preferable if $N_R < \bar{N}_R$, and the LMMSE-type estimator is preferable otherwise. Note that \bar{N}_R is different for UL and DL covariance estimation. It can be obtained by equating the SINR expressions for the LMMSE-type and element-wise LMMSE-type channel estimation techniques (for UL and DL) and solving the corresponding equation for N_R . After some straight forward algebra, \bar{N}_R can be obtained in the form

$$\bar{N}_R = \frac{fc - ah}{ag - fb} \quad (44)$$

where b , c , g and h are given at the bottom of the next page along with the following parameters

$$\begin{aligned} a &= \left(\frac{N_Q}{N_Q - M} \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) \right)^2 \\ f &= \left(\frac{N_Q}{(N_Q - 1)} \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) \right)^2 \\ \bar{\mathbf{S}}_s &= \text{diag}(\bar{\mathbf{R}}_s); \quad \bar{\mathbf{R}}_s = \begin{cases} \mathbf{R}_s, & \text{for UL} \\ \mathbf{R}_s^{(dl)}, & \text{for DL} \end{cases} \\ d &= \begin{cases} 0, & \text{for UL} \\ \frac{1}{\lambda}, & \text{for DL.} \end{cases} \end{aligned}$$

Note that \bar{N}_R is a function of N_Q which can take any real value. Thus, if \bar{N}_R is negative for some value of N_Q ,

it means, for that particular choice of N_Q , there is no valid N_R that makes the LMMSE-type channel estimation preferable. Consequently, using (44), we can also compute a threshold for N_Q below which element-wise LMMSE-type channel estimation is always preferred. However, deriving a theoretical expression for such a threshold is extremely difficult. It can be easily computed numerically.

Therefore, the closed-form expressions for average UL and DL SE, for the LMMSE-type and element-wise LMMSE-type channel estimation methods serve as tools for choosing different design parameters, and also as a tool for choosing a preferred channel estimation technique. In practice, with approximate models of the covariance matrix of an individual user in a massive MIMO system, the derived expressions for average SE enables us to choose these parameters for the desired UL and DL SE values.

In what follows, we validate the derived theoretical SE expressions with simulated SE obtained by averaging over multiple realizations of random covariance estimation matrices. Then, we compare the theoretical SE expressions with the SE expressions that correspond to known covariance case. Finally, we also depict the impact of N_R on the SE.

V. SIMULATION RESULTS

We consider a massive MIMO system with $L = 7$ cells, each comprising a BS with $M = 100$ antennas and $K = 10$ users. The BSs are at a distance of 300 m apart from each other, and the users are uniformly spaced at a distance of 120 m from the BS in their cells. The users that reuse the same pilot in different cells are at the same position relative to the corresponding BSs. The angular spread of the channel cluster is assumed to be 20° , within which the received paths from a user are assumed to be uniformly distributed. We consider a 3GPP urban macro (UMa) [20] scenario with a non-line-of-sight (N-LOS) channel for simulating the path loss model. The mean path loss of the received signal from a

$$\begin{aligned} & \mathbb{E}_W \{ \text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s) \} \\ &= \text{tr}(\mathbf{W}_{ju}^H \mathbf{Q}_{ju} \mathbf{W}_{ju}^H \mathbf{R}_s) + \left\{ \kappa_3 \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{Q}_{ju} \bar{\mathbf{W}}_{ju}^H \mathbf{R}_s) - \text{tr}(\mathbf{W}_{ju}^H \mathbf{Q}_{ju} \mathbf{W}_{ju}^H \mathbf{R}_s) \right. \\ & \quad + \frac{\alpha_R^2 \kappa_3}{2N_R} \text{tr}(\mathbf{P}_{ju}^{-1} \mathbf{Q}_{ju} \mathbf{P}_{ju}^{-1} \{ \mathbf{R}_s \circ \mathbf{Q}_{ju} \circ \mathbf{Q}_{ju} \} + \mathbf{P}_{ju}^{-1} \mathbf{Q}_{ju} \mathbf{P}_{ju}^{-1} \{ \mathbf{R}_s \circ \mathbf{R}_{jju} \circ \mathbf{R}_{jju} \}) + \kappa_4 \text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{P}_{ju} \bar{\mathbf{W}}_{ju}^H \mathbf{S}_s) \\ & \quad \left. + \frac{\alpha_R^2 \kappa_4}{2N_R} \text{tr}(\mathbf{S}_s \mathbf{P}_{ju}) + \frac{\alpha_R^2 \kappa_4}{2N_R} \text{tr}(\mathbf{W}_{ju}^H \mathbf{S}_s \mathbf{W}_{ju}) \right\} \quad (36) \end{aligned}$$

$$\begin{aligned} & \mathbb{E}_W \{ |\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2 \} \\ &= |\text{tr}(\mathbf{W}_{ju}^H \mathbf{S}_{jlu})|^2 + \left\{ \kappa_3 |\text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{S}_{jlu})|^2 - |\text{tr}(\mathbf{W}_{ju}^H \mathbf{S}_{jlu})|^2 + \frac{\alpha_R^2 \kappa_3}{2N_R} \sum_{p=1}^M \sum_{q=1}^M |\mathbf{W}_{lu}(\mathbf{Q}_{ju} \circ \mathbf{Q}_{ju}) \mathbf{W}_{lu}|_{pq} \right. \\ & \quad \left. + \frac{\alpha_R^2 \kappa_3}{2N_R} \sum_{p=1}^M \sum_{q=1}^M |\mathbf{W}_{lu}(\mathbf{R}_{jju} \circ \mathbf{R}_{jju}) \mathbf{W}_{lu}|_{pq} + \kappa_4 \text{tr}(\bar{\mathbf{W}}_{ju}^2 \mathbf{S}_{jlu}^2) + \frac{\alpha_R^2 \kappa_4}{2N_R} \text{tr}(\mathbf{W}_{lu}^2 \mathbf{P}_{ju}^2) + \frac{\alpha_R^2 \kappa_4}{2N_R} \text{tr}(\mathbf{W}_{lu}^2 \mathbf{S}_{jju}^2) \right\} \quad (37) \end{aligned}$$

$$\gamma_{ju}^{(dl)} = \frac{|\mathbb{E}_W \{ \text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) \}|^2}{\mathbb{E}_W \{ \text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{Q}_{ju} \hat{\mathbf{W}}_{ju}^H \mathbf{R}_s^{(dl)}) \} + \sum_{l=1}^L \mathbb{E}_W \{ |\text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2 \} - |\mathbb{E}_W \{ \text{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju}) \}|^2 + \frac{1}{\lambda}} \quad (38)$$

user that is at a distance d (in m) from the BS is given as $PL(f, d) = 32.4 + 20 \log_{10}(f/1 \text{ GHz}) + 30 \log_{10}(d_{3D}/1 \text{ m})$, where $d_{3D} = \sqrt{d^2 + (h_{BS} - h_{UT})^2}$ m, f is the carrier frequency, h_{BS} is the height of a BS in m, and h_{UT} is the height of a UE in m. Therefore, the mean received SNR, in dB, is given by $SNR = P_T - PL - 10 \log_{10}(kT_0B) - NF$, where P_T is the transmit power, k is the Boltzmann constant, $T_0 = 290$ K is the nominal temperature, B is the signal bandwidth, and NF is the noise figure in dB. In this setup, we consider $f = 3.4$ GHz, $P_T = 6$ dBm, $B = 40$ MHz, $NF = 10$ dB, $h_{BS} = 25$ m, and $h_{UT} = 1.5$ m which results in the mean SNR of the received signal from a user that is at a distance d from the BS to be given by $46.93 - 30 \log_{10} d_{3D}$.

The number of symbols that are dedicated for UL transmission within each coherence block is chosen to be $C_u = 100$ symbols. We choose the number of symbols used for the ChEst (and also the CovEst) pilot to be $P = 10$. Second-order statistics of the channel are assumed to be constant for $\tau_s = 25000$ coherence blocks, and the UL transmit power is $\mu = 1$, and the DL transmit power is $\lambda = 10$. Additionally, we choose $\alpha_R = 0.95$, and $\mathbf{R}_b = \mathbf{I}$. Sample averaging for all the expectation terms is computed using 2000 trials for different values of $N_R = \{125, 250, 500, 1000, 2000, 4000, 8000\}$.

A. Uplink Spectral Efficiency

For this simulation example, we consider the UL SE expressions that correspond to the two channel estimation techniques: LMMSE-type channel estimation and the element-wise LMMSE-type channel estimation, each in combination with two beamforming techniques, which are MRC and ZF combining.⁵ In Fig. 3, we plot the SE as a function of N_R for the two aforementioned channel estimation techniques and beamforming techniques. Fig. 3(a) depicts the SE values for $N_Q = 125$ and Fig. 3(b) shows SE values for $N_Q = 4000$. In both the subplots, we present SE values corresponding to known covariance matrices and theoretical SE values (only for the MRC combining case) as well as simulated SE values

⁵In practice, ZF is a good choice for beamforming [21].

corresponding to the two channel estimation techniques that use the estimated covariance matrices.

In Fig. 3, it can be noticed that the theoretical SE, corresponding to LMMSE-type channel estimation and MRC combining, initially rises with N_R to approach the SE that corresponds to LMMSE channel estimation, followed by a drop in the theoretical SE at $N_R = 8000$. In contrast, the theoretical SE, corresponding to element-wise LMMSE-type channel estimation and MRC combining, approaches the SE corresponding to element-wise LMMSE channel estimation for N_R value as low as 125 and reaches its maximum at $N_R = 500$. Then, the theoretical SE reduces linearly with a further increase in N_R as the pilot overhead increases. Moreover, the simulated SEs match the theoretical values for both the channel estimation techniques tested, thereby validating the derivations presented in the paper.

The initial rise of the theoretical SEs is due to the improvement in the covariance estimates caused by the increase in the number of samples for estimation. However, a further increase in N_R results in a drop in UL SEs due to the pre-log factor. Despite the improvement in estimation quality of the covariance matrices, the SEs drop because of the consumption of UL resources for the additional CovEst pilots. This validates the theoretical analysis done in Section IV. Moreover, it should be noted that the LMMSE should always perform better than element-wise LMMSE as the correlation between antenna elements' channels are ignored in element-wise LMMSE. However, due to imperfect covariance information, element-wise LMMSE-type is not necessarily better than LMMSE-type. Specifically, element-wise LMMSE-type is expected to outperform the LMMSE-type when $N_R \leq \bar{N}_R$.

It can be seen from Fig. 3(a) and Fig. 3(b) that using element-wise LMMSE channel estimation instead of LMMSE channel estimation leads to a drop in SE. However, it is evident that the element-wise LMMSE-type channel estimation completely outperforms the LMMSE-type channel estimation for all the N_R values and for $N_Q = 125$. It can also be noted that even for $N_Q = 4000$, the element-wise LMMSE-type channel estimation outperforms the LMMSE-type channel estimation

$$\begin{aligned}
 b &= \kappa_1 \text{tr}(\bar{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \bar{\mathbf{W}}_{ju}^H \bar{\mathbf{R}}_s) + \sum_{l=1}^L \left\{ \kappa_2 |\text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jlu})|^2 + \frac{\kappa_1}{N_Q} \text{tr}(\bar{\mathbf{W}}_{ju}^H \bar{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \mathbf{W}_{lu}^H \mathbf{W}_{lu} \mathbf{Q}_{ju}) \right\} - a + d \\
 c &= \frac{\alpha_R^2 \kappa_1}{2} \{ M \text{tr}(\bar{\mathbf{R}}_s \mathbf{Q}_{ju}) + \text{tr}(\mathbf{W}_{ju}) \text{tr}(\bar{\mathbf{R}}_s \mathbf{R}_{jju}) \} + \frac{\alpha_R^2 \kappa_2}{2} \sum_{l=1}^L \{ \text{tr}(\mathbf{W}_{lu} \mathbf{Q}_{ju} \mathbf{W}_{lu}^H \mathbf{Q}_{ju}) + \text{tr}(\mathbf{W}_{lu} \mathbf{R}_{jju} \mathbf{W}_{lu}^H \mathbf{R}_{jju}) \} \\
 &\quad + \frac{\alpha_R^2 \kappa_1}{2N_Q} \sum_{l=1}^L \{ M \text{tr}(\mathbf{W}_{jlu}^2 \mathbf{Q}_{ju}^2) + \text{tr}(\mathbf{W}_{ju}) \text{tr}(\mathbf{W}_{jlu}^2 \mathbf{Q}_{ju} \mathbf{R}_{jju}) \} \\
 g &= \kappa_3 \left\{ \text{tr}(\bar{\mathbf{W}}_{ju} \mathbf{Q}_{ju} \bar{\mathbf{W}}_{ju}^H \bar{\mathbf{R}}_s) + \sum_{l=1}^L |\text{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{S}_{jlu})|^2 \right\} + \kappa_4 \left\{ \text{tr}(\bar{\mathbf{W}}_{ju} \mathbf{P}_{ju} \bar{\mathbf{W}}_{ju}^H \bar{\mathbf{S}}_s) + \sum_{l=1}^L \text{tr}(\bar{\mathbf{W}}_{ju}^2 \mathbf{S}_{jlu}^2) \right\} - f + d \\
 h &= \frac{\alpha_R^2 \kappa_3}{2} \text{tr}(\mathbf{P}_{ju}^{-1} \mathbf{Q}_{ju} \mathbf{P}_{ju}^{-1} \{ \bar{\mathbf{R}}_s \circ \mathbf{Q}_{ju} \circ \mathbf{Q}_{ju} \} + \mathbf{P}_{ju}^{-1} \mathbf{Q}_{ju} \mathbf{P}_{ju}^{-1} \{ \bar{\mathbf{R}}_s \circ \mathbf{R}_{jju} \circ \mathbf{R}_{jju} \}) + \frac{\alpha_R^2 \kappa_4}{2} \{ \text{tr}(\bar{\mathbf{S}}_s \mathbf{P}_{ju}) + \text{tr}(\mathbf{W}_{ju} \bar{\mathbf{S}}_s \mathbf{S}_{jju}) \\
 &\quad + \sum_{l=1}^L \text{tr}(\mathbf{W}_{lu}^2 \mathbf{P}_{ju}^2) + \sum_{l=1}^L \text{tr}(\mathbf{W}_{lu}^2 \mathbf{S}_{jju}^2) \} + \frac{\alpha_R^2 \kappa_3}{2} \sum_{p=1}^M \sum_{q=1}^M \{ [\mathbf{W}_{lu}(\mathbf{Q}_{ju} \circ \mathbf{Q}_{ju}) \mathbf{W}_{lu}]_{pq} + [\mathbf{W}_{lu}(\mathbf{R}_{jju} \circ \mathbf{R}_{jju}) \mathbf{W}_{lu}]_{pq} \}
 \end{aligned}$$

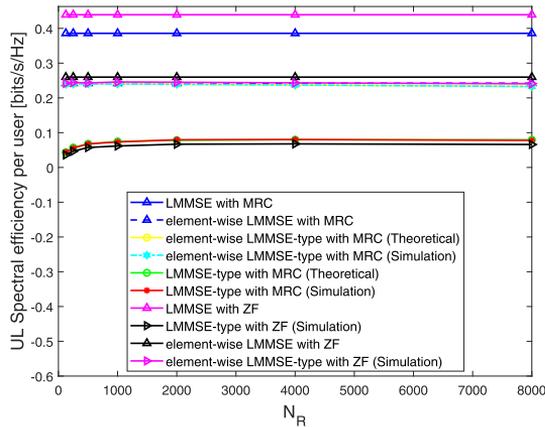
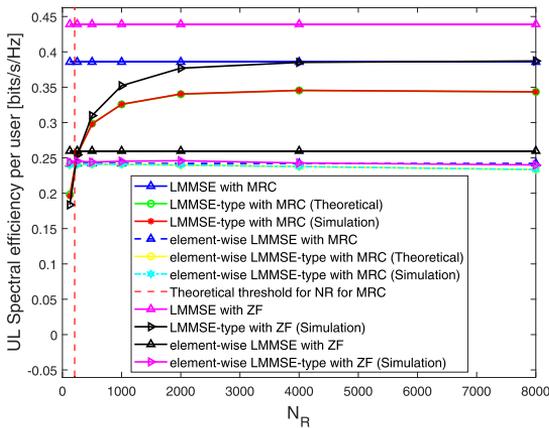
(a) SE vs N_R with $N_Q = 125$.(b) SE vs N_R with $N_Q = 4000$.

Fig. 3. UL SE for different channel estimation techniques.

for $N_R = 125$. Moreover, for $N_Q = 4000$, \bar{N}_R given in Section IV matches exactly with the N_R value for which the LMMSE-type and element-wise LMMSE-type channel estimations have the same performance. Therefore, the minimum SE guaranteed for a massive MIMO system with imperfect covariance information is the SE provided by the element-wise LMMSE channel estimator.⁶ This SE can be achieved by using element-wise LMMSE-type channel estimation with very low values of N_R and N_Q , and with low computational complexity. From simulations, we also observe the threshold value for N_Q to be 263, such that for $N_Q < 263$, element-wise LMMSE-type channel estimation always outperforms LMMSE-type channel estimation.

In Fig. 3, it can also be seen that the SE simulation curves corresponding to ZF combining behave similarly to the case of MRC combining. ZF combining performs well only for

⁶Note that the objective is to have N_R and N_Q as low as possible for guaranteeing a desired SE.

the larger number of pilots ($N_Q = 4000$ and $N_R \geq 500$) and needs additional computational complexity, but it does not significantly improve the performance of element-wise LMMSE channel estimation and it gives only marginally better performance than that corresponding to MRC combining. Moreover, for $N_Q = 4000$, the SE curve for LMMSE-type channel estimation crosses the SE for element-wise LMMSE-type channel estimation closer to the \bar{N}_R value theoretically computed for MRC combining (44). For large N_R and N_Q values, ZF combining outperforms MRC combining. This is due to better covariance estimates. Therefore, the SE expressions derived in this paper serve as conservative bounds for an achievable spectral efficiency of the system considered.

B. Downlink Spectral Efficiency

Similar to the UL example, in this simulation example, we consider the DL SE expressions that correspond to the two channel estimation techniques: LMMSE-type channel estimation and the element-wise LMMSE-type channel estimation, each in combination with two beamforming techniques, which are matched filter precoding and ZF precoding. In Fig. 4, we plot the SE as a function of N_R for the two aforementioned channel estimation techniques. Fig. 4(a) depicts the SE values for $N_Q = 125$, and Fig. 4(b) shows SE values for $N_Q = 4000$. We perform a study on these plots similar to the study done in Subsection V-A.

It can be observed from Fig. 4 that the DL SE plots are similar to the plots in Subsection V-A. However, unlike in UL SE, an increase in N_R does not result in a drop in SE as there is no pilot overhead in DL. The simulated SEs match the theoretical values for both the channel estimation techniques used, thereby validating the derivations presented in the paper. Moreover, for $N_Q = 4000$, \bar{N}_R given in Section IV matches exactly with the N_R value for which LMMSE-type and element-wise LMMSE-type channel estimations have the same performance. From Fig. 4(a) and Fig. 4(b), the minimum DL SE guaranteed for a massive MIMO system with imperfect covariance information is the SE provided by element-wise LMMSE channel estimator. This SE can be achieved by using element-wise LMMSE-type channel estimation with very low values of N_R and N_Q , with low computational complexity. From simulations, we also compute the threshold value for N_Q to be 272, such that for $N_Q < 272$, element-wise LMMSE-type channel estimation always outperforms LMMSE-type channel estimation.

It can also be noticed from Fig. 4 that the SE simulation curves corresponding to ZF precoding behave similarly to the case of matched filter precoding. ZF precoding performs well only for the larger number of pilots ($N_Q = 4000$ and $N_R \geq 500$) and needs additional computational complexity, but it does not significantly improve the performance corresponding to element-wise LMMSE channel estimation, and it gives only marginally better performance than that corresponding to matched filter precoding. Moreover, for $N_Q = 4000$, the SE curve for LMMSE-type channel estimation crosses the SE for element-wise LMMSE-type channel estimation closer to the \bar{N}_R value theoretically computed for matched filter precoding (44). For large N_R and N_Q values, ZF precoding

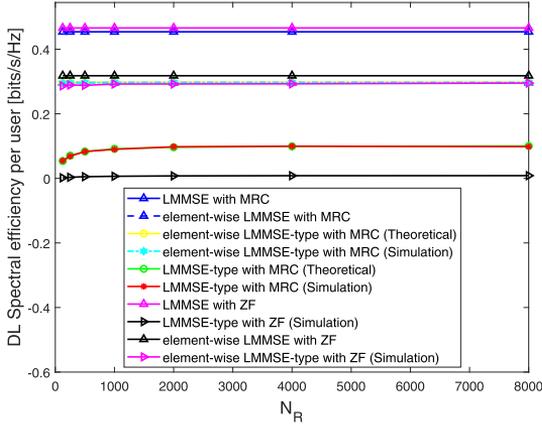
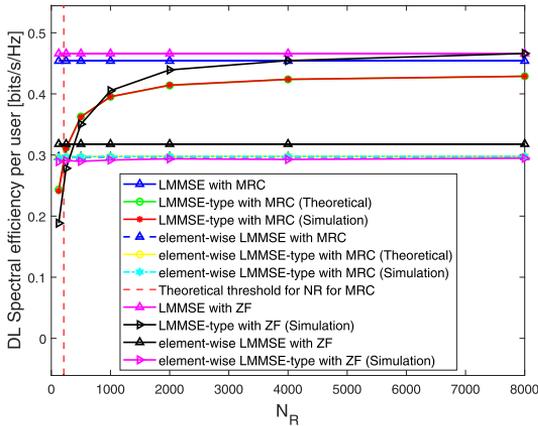

 (a) SE vs N_R with $N_Q = 125$.

 (b) SE vs N_R with $N_Q = 4000$.

Fig. 4. DL SE for different channel estimation techniques.

outperforms the matched filter precoding. This is due to better covariance estimates. Therefore, the SE expressions derived in this paper serve as conservative bounds for an achievable spectral efficiency of the system considered.

VI. CONCLUSION

We have derived closed-form expressions for average UL and DL SEs of a massive MIMO system that implements MRC and matched filter precoder, respectively, as a function of N_R and N_Q , which represent the UL pilot overhead. These combiners use channel estimates that utilize estimated covariance matrices in addition to channel observations. The LMMSE-type and element-wise LMMSE-type channel estimates have been considered. Using theoretical analysis of these closed-form expressions as well as simulation results, we have demonstrated the impact of different values of N_R and N_Q on SEs of a user in a massive MIMO system,

thereby presenting the closed-form expressions as the tools for solving the problem of choosing these parameters optimally. Based on numerical study, we have demonstrated that the ZF beamforming does not significantly improve the SE for the case of element-wise LMMSE-type channel estimation. It is useful for the case of LMMSE-type channel estimation but at the cost of large pilot overhead and computational complexity. Furthermore, we have shown that the choice of pilot overhead made based on the LS beamforming serves as a more conservative result than for the ZF beamforming case, but accurate and very useful estimate of the pilot overhead. Finally, we have shown that the element-wise LMMSE-type channel estimator with very low N_R and N_Q and with simple LS combiner provides the minimum SE guarantee with low computational complexity.

APPENDIX A

PROOF OF LEMMA 1

Let us start with a proof of (17). Let the rank of the covariance matrix of \mathbf{h} , \mathbf{R} , be K . Then, we denote $\mathbf{\Lambda} \in \mathbb{R}^{K \times K}$ is a diagonal matrix containing positive eigenvalues of \mathbf{R} and $\mathbf{U} \in \mathbb{R}^{M \times K}$ is a matrix containing K eigenvectors corresponding to eigenvalues. Now, let us also define $\mathbf{B} \triangleq \mathbf{U}\mathbf{\Lambda}^{1/2} \in \mathbb{C}^{M \times K}$. Then, there exists a unique $\mathbf{g} \in \mathbb{C}^K$ such that $\mathbf{h} = \mathbf{B}\mathbf{g}$ and $\mathbb{E}\{\mathbf{g}\mathbf{g}^H\} = \mathbf{I}$. Therefore, we have $\mathbb{E}\{\mathbf{h}\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\} = \mathbf{B}\mathbb{E}\{\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H\}\mathbf{B}^H$ where $\tilde{\mathbf{A}} \triangleq \mathbf{B}^H\mathbf{A}\mathbf{B}$. However, since \mathbf{g} is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I})$, the term $\mathbb{E}\{\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H\}$ can be evaluated as follows

$$\begin{aligned} \mathbb{E}\{[\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H]_{ij}\} &= \sum_{p=1}^K \sum_{q=1}^K \mathbb{E}\{[\mathbf{g}]_i [\mathbf{g}]_p^* [\mathbf{g}]_q [\mathbf{g}]_j^*\} [\tilde{\mathbf{A}}]_{pq} \\ &= \begin{cases} [\tilde{\mathbf{A}}]_{ij} & \text{if } i \neq j \\ [\tilde{\mathbf{A}}]_{ii} + \text{tr}(\tilde{\mathbf{A}}) & \text{otherwise} \end{cases} \end{aligned}$$

and $\mathbb{E}\{\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H\} = \tilde{\mathbf{A}} + \text{Itr}(\tilde{\mathbf{A}})$. Therefore, $\mathbb{E}\{\mathbf{h}\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\} = \mathbf{R}\mathbf{A}\mathbf{R} + \mathbf{R}\text{tr}(\mathbf{A}\mathbf{R})$.

Proof of (18) is as follows. We first compute that $\mathbb{E}\{|\mathbf{h}^H\mathbf{A}\mathbf{h}|^2\} = \mathbb{E}\{\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\} = \mathbb{E}\{\text{tr}(\mathbf{A}\mathbf{h}\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H)\}$. Using (17), we have $\mathbb{E}\{|\mathbf{h}^H\mathbf{A}\mathbf{h}|^2\} = |\text{tr}(\mathbf{A}^H\mathbf{R})|^2 + \text{tr}(\mathbf{A}\mathbf{R}\mathbf{A}^H\mathbf{R})$.

APPENDIX B

PROOF OF LEMMA 2

Proof of (19) and (20) is given in [22].

Using the eigenvalue decomposition of $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, let us define $\tilde{\mathbf{X}} \triangleq \mathbf{U}^H\mathbf{X}\mathbf{U}$. It should be noted that $\tilde{\mathbf{X}}$ is distributed as $\mathcal{W}(N, \mathbf{I})$. Then, (21) can be proved as follows. First, we compute the following expectation term.

$$\mathbb{E}\{\text{tr}(\mathbf{X}^{-2}\mathbf{C})\} = \mathbb{E}\{\text{tr}(\tilde{\mathbf{X}}^{-2}\mathbf{\Lambda})\} = \sum_{i=1}^M \mathbb{E}\{[\tilde{\mathbf{X}}^{-2}]_{ii}\} [\mathbf{\Lambda}]_{ii}$$

But from (20), we have

$$\begin{aligned} \mathbb{E}\{\text{tr}(\mathbf{X}^{-2}\mathbf{C})\} &= \sum_{i=1}^M \frac{N}{(N-M)^3 - (N-M)} [\mathbf{\Lambda}]_{ii} \\ &= \frac{N}{(N-M)^3 - (N-M)} \text{tr}(\mathbf{C}) \end{aligned}$$

For (21), we expand $\mathbb{E}\{|\text{tr}(\mathbf{X}^{-1}\mathbf{A})|^2\}$ using (20) as follows.

$$\begin{aligned} & \mathbb{E}\{|\text{tr}(\mathbf{X}^{-1}\mathbf{A})|^2\} \\ &= \sum_{p=1}^M \sum_{q=1}^M \sum_{r=1}^M \sum_{s=1}^M \mathbb{E}\{[\mathbf{X}^{-1}]_{pp}[\mathbf{X}^{-1}]_{ss}[\mathbf{A}]_{pp}[\mathbf{A}^H]_{ss}\} \\ &= \sum_{p=1}^M \mathbb{E}\{[\mathbf{X}^{-1}]_{pp}[\mathbf{X}^{-1}]_{pp}[\mathbf{A}]_{pp}[\mathbf{A}^H]_{pp}\} \\ &+ \sum_{p=1}^M \sum_{s=1, s \neq p}^M \mathbb{E}\{[\mathbf{X}^{-1}]_{pp}[\mathbf{X}^{-1}]_{ss}[\mathbf{A}]_{pp}[\mathbf{A}^H]_{ss}\} \\ &+ \sum_{p=1}^M \sum_{s=1, s \neq p}^M \mathbb{E}\{[\mathbf{X}^{-1}]_{ps}[\mathbf{X}^{-1}]_{sp}[\mathbf{A}]_{sp}[\mathbf{A}^H]_{ps}\} \end{aligned}$$

Using (20), the above equation can be further simplified to (22).

APPENDIX C PROOF OF LEMMA 3

Let us define a pair of mutually independent random vectors as follows.

$$\mathbf{g}_{jju}^{(1)}[n] \triangleq \hat{\mathbf{h}}_{jju}^{(1)}[n] - \mathbf{h}_{jju}, \quad \mathbf{g}_{jju}^{(2)}[n] \triangleq \hat{\mathbf{h}}_{jju}^{(2)}[n] - \mathbf{h}_{jju}$$

The covariance matrices for $\mathbf{g}^{(1)}[n]$ and $\mathbf{g}^{(2)}[n]$ are identically equal to $\mathbf{Q}_{ju} - \mathbf{R}_{jju}$. Additionally, we also define mutually independent set of matrices

$$\check{\mathbf{R}}_{jju}[n] \triangleq \hat{\mathbf{h}}_{jju}^{(1)}[n](\hat{\mathbf{h}}_{jju}^{(2)}[n])^H + \hat{\mathbf{h}}_{jju}^{(2)}[n](\hat{\mathbf{h}}_{jju}^{(1)}[n])^H$$

for all $n \in \{1 \dots N_R\}$ such that $\check{\mathbf{R}}_{jju} = \frac{1}{N_R} \sum_{n=1}^N \check{\mathbf{R}}_{jju}[n]$ by definition (i.e., (10)).

Using the definition of $\mathbf{g}_{jju}^{(1)}[n]$ and $\mathbf{g}_{jju}^{(2)}[n]$, and also using Lemma 1, it can be shown that, for all $n = \{1 \dots N_R\}$, we have

$$\begin{aligned} \mathbb{E}\{\check{\mathbf{R}}_{jju}[n]\mathbf{A}\check{\mathbf{R}}_{jju}[n]\} &= \mathbf{R}_{jju}\mathbf{A}\mathbf{R}_{jju} + \frac{1}{2}\mathbf{Q}_{ju}\text{tr}(\mathbf{A}\mathbf{Q}_{ju}) \\ &+ \frac{1}{2}\mathbf{R}_{jju}\text{tr}(\mathbf{A}\mathbf{R}_{jju}) \end{aligned} \quad (45)$$

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\check{\mathbf{R}}_{jju}[n]\mathbf{A})|^2\} &= |\text{tr}(\mathbf{R}_{jju}\mathbf{A})|^2 + \frac{1}{2}\text{tr}(\mathbf{A}\mathbf{Q}_{ju}\mathbf{A}^H\mathbf{Q}_{ju}) \\ &+ \frac{1}{2}\text{tr}(\mathbf{A}\mathbf{R}_{jju}\mathbf{A}^H\mathbf{R}_{jju}). \end{aligned} \quad (46)$$

Finally, along with the equation $\check{\mathbf{R}}_{jju} = \frac{1}{N_R} \sum_{n=1}^N \check{\mathbf{R}}_{jju}[n]$, (45) and (46) will result in (23) and (24), respectively.

APPENDIX D PROOF OF LEMMA 5

Since $\mathbf{Y} = \mathbf{Z}/2$, and the elements of the diagonal matrix \mathbf{Z} are χ^2 distributed with $2N$ degrees of freedom, we have $\mathbb{E}\{[\mathbf{Y}^{-1}]_{pp}\} = 2\mathbb{E}\{[\mathbf{Z}^{-1}]_{pp}\} = 1/(N-1)$ and $\mathbb{E}\{[\mathbf{Y}^{-1}]_{pp}^2\} = 4\mathbb{E}\{[\mathbf{Z}^{-1}]_{pp}^2\} = 1/(N-1)(N-2)$.

Using the above results, (31) can be derived as follows

$$\begin{aligned} & \mathbb{E}\{\text{tr}(\mathbf{Y}^{-1}\mathbf{A}_1\mathbf{Y}^{-1}\mathbf{A}_2)\} \\ &= \left(\frac{1}{N-1}\right)^2 \sum_{p=1}^M \sum_{q \neq p}^M [\mathbf{A}_1]_{pq}[\mathbf{A}_2]_{qp} \end{aligned}$$

$$\begin{aligned} & + \frac{1}{(N-1)(N-2)} \sum_{p=1}^M [\mathbf{A}_1]_{pp}[\mathbf{A}_2]_{pp} \\ &= \tau_1 \text{tr}(\mathbf{A}_1\mathbf{A}_2) + \tau_2 \text{tr}(\mathbf{A}_{1d}\mathbf{A}_{2d}) \end{aligned}$$

where $\tau_1 \triangleq 1/(N-1)^2$, $\tau_2 \triangleq 1/((N-1)^2(N-2))$, $\mathbf{A}_{1d} \triangleq \text{diag}(\mathbf{A}_1)$, and $\mathbf{A}_{2d} \triangleq \text{diag}(\mathbf{A}_2)$.

In what follows, proof of (32) is presented

$$\begin{aligned} \mathbb{E}\{|\text{tr}(\mathbf{Y}^{-1}\mathbf{A})|^2\} &= \frac{1}{(N-1)^2} \sum_{p=1}^M \sum_{q \neq p}^M [\mathbf{A}]_{pp}[\mathbf{A}]_{qq}^* \\ &+ \frac{1}{(N-1)(N-2)} \sum_{p=1}^M |[\mathbf{A}]_{pp}|^2 \\ &= \tau_1 |\text{tr}(\mathbf{A})|^2 + \tau_2 \text{tr}(\mathbf{A}_d^H \mathbf{A}_d) \end{aligned}$$

where $\mathbf{A}_d \triangleq \text{diag}(\mathbf{A})$.

APPENDIX E PROOF OF LEMMA 6

Let us define a pair of mutually independent random vectors as follows.

$$\mathbf{g}_{jju}^{(1)}[n] \triangleq \hat{\mathbf{h}}_{jju}^{(1)}[n] - \mathbf{h}_{jju}, \quad \mathbf{g}_{jju}^{(2)}[n] \triangleq \hat{\mathbf{h}}_{jju}^{(2)}[n] - \mathbf{h}_{jju}$$

The covariance matrices for $\mathbf{g}_{jju}^{(1)}[n]$ and $\mathbf{g}_{jju}^{(2)}[n]$ are identically equal to $\mathbf{Q}_{ju} - \mathbf{R}_{jju}$. Additionally, we also define mutually independent set of matrices as $\check{\mathbf{S}}_{jju}[n] \triangleq \text{diag}(\hat{\mathbf{h}}_{jju}^{(1)}[n](\hat{\mathbf{h}}_{jju}^{(2)}[n])^H + \hat{\mathbf{h}}_{jju}^{(2)}[n](\hat{\mathbf{h}}_{jju}^{(1)}[n])^H)$ for all $n \in \{1 \dots N_R\}$ such that $\check{\mathbf{S}}_{jju} = \frac{1}{N} \sum_{n=1}^N \check{\mathbf{S}}_{jju}[n]$ by definition (i.e., (12)).

Using the definitions of $\mathbf{g}_{jju}^{(1)}[n]$ and $\mathbf{g}_{jju}^{(2)}[n]$ together with Lemma 1 (for scalar case), and Lemma 4, it can be shown that

$$\begin{aligned} & \mathbb{E}\{\{\check{\mathbf{S}}_{jju}\}_{pp}[\check{\mathbf{S}}_{jju}\]_{qq}\} \\ &= \mathbb{E}\{|\mathbf{h}_{jju}|_p|^2 |\mathbf{h}_{jju}|_q|^2\} \\ &+ \frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{Q}_{ju} - \mathbf{R}_{jju}]_{qp} + \frac{1}{2}[\mathbf{Q}_{ju} - \mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{qp} \\ &+ \frac{1}{2}[\mathbf{Q}_{ju} - \mathbf{R}_{jju}]_{pq}[\mathbf{Q}_{ju} - \mathbf{R}_{jju}]_{qp} \\ &= [\mathbf{S}_{jju}]_{pp}[\mathbf{S}_{jju}]_{qq} + \frac{1}{2}[\mathbf{Q}_{jju}]_{pq}[\mathbf{Q}_{jju}]_{pq} \\ &+ \frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{pq}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} & \mathbb{E}\{\{\check{\mathbf{S}}_{jju}\mathbf{A}\check{\mathbf{S}}_{jju}\}_{pq}\} \\ &= [\mathbf{A}]_{pq} \left\{ [\mathbf{S}_{jju}]_{pp}[\mathbf{S}_{jju}]_{qq} + \frac{1}{2}[\mathbf{Q}_{jju}]_{pq}[\mathbf{Q}_{jju}]_{pq} \right. \\ &\quad \left. + \frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{pq} \right\} \quad (47) \\ & \mathbb{E}\{|\text{tr}(\check{\mathbf{S}}_{jju}\mathbf{D})|^2\} \\ &= \sum_{p=1}^M \sum_{q=1}^M \left\{ [\mathbf{S}_{jju}]_{pp}[\mathbf{S}_{jju}]_{qq} + \frac{1}{2}[\mathbf{Q}_{ju}]_{pq}[\mathbf{Q}_{ju}]_{pq} \right. \\ &\quad \left. + \frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{pq} \right\} [\mathbf{D}]_{pp}[\mathbf{D}]_{qq} \end{aligned}$$

$$\begin{aligned}
&= |\text{tr}(\mathbf{S}_{jju}\mathbf{D})|^2 + \frac{1}{2} \sum_{p=1}^M \sum_{q=1}^M [\mathbf{D}(\mathbf{Q}_{ju} \circ \mathbf{Q}_{ju})\mathbf{D}]_{pq} \\
&+ \frac{1}{2} \sum_{p=1}^M \sum_{q=1}^M [\mathbf{D}(\mathbf{R}_{jju} \circ \mathbf{R}_{jju})\mathbf{D}]_{pq}. \quad (48)
\end{aligned}$$

Finally, along with the equation $\tilde{\mathbf{S}}_{jju} = \frac{1}{N} \sum_{n=1}^N \tilde{\mathbf{S}}_{jju}[n]$, (47) and (48) will result in (33) and (34), respectively.

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [4] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [5] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [6] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.
- [7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [8] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [9] A. Pitarokoulis, E. Björnson, and E. G. Larsson, "On the effect of imperfect timing synchronization on pilot contamination," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
- [10] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.
- [11] S. Haghighatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8316–8332, Dec. 2017.
- [12] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.
- [13] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.
- [14] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, Jun. 2018.
- [15] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. 50th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 974–978.
- [16] K. Upadhyaya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546–550, Apr. 2018.
- [17] A. K. Kocharlakota, K. Upadhyaya, and S. A. Vorobyov, "On achievable rates for massive MIMO system with imperfect channel covariance information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4504–4508.
- [18] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.
- [19] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *Proc. 5th Meeting COST*, Lisbon, Portugal, vol. 273, Sep. 2002, pp. 1–10.
- [20] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document 3GPP, (TR) 38.901, version 16.1.0, Dec. 2019. [Online]. Available: http://www.3gpp.org/specs/archive/38_series/38.901/38901-g10.zip

- [21] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.
- [22] D. Maiwald and D. Kraus, "On moments of complex Wishart and complex inverse Wishart distributed matrices," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, vol. 5, Apr. 1997, pp. 3817–3820.



Atchutaram K. Kocharlakota (Student Member, IEEE) received the M.Tech. degree from IIT Madras, India, in 2011. He is currently pursuing the D.Sc.(Tech.) degree from Aalto University, Finland. Prior to his Ph.D., he was part of the Research and Development Team, Intel Mobile Communications, India, from 2011 to 2017. His research interests include massive MIMO, cell-free massive MIMO, and intelligent reflecting surface aided communications.



Karthik Upadhyaya (Member, IEEE) received the M.Tech. degree from IIT Madras, India, in 2011, and the D.Sc.(Tech.) degree from Aalto University, Finland, in 2018. Prior to his Ph.D., he was a Member of Technical Staff with Saankhya Labs, India, from 2011 to 2013, and a Research Assistant with Indian Institute of Science, India, from 2013 to 2014. He was a Visiting Researcher with the Wireless Networking and Communications Group, The University of Texas at Austin, in 2017. He is currently with Nokia Bell Labs, Espoo, Finland. His research interests include massive MIMO and physical layer security.



Sergiy A. Vorobyov (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in systems and control from Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, in 1994 and 1997, respectively.

He is currently a Professor with the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland. He has been previously with the University of Alberta, AB, Canada, as an Assistant Professor and an Associate Professor, and then as a Full Professor. Since his graduation, he has been holding various research and faculty positions with Kharkiv National University of Radio Electronics; the Institute of Physical and Chemical Research, Japan; McMaster University, Canada; the University of Duisburg-Essen, Germany; Darmstadt University of Technology, Germany; and the Joint Research Institute between Heriot-Watt University and The University of Edinburgh, U.K. His research interests include optimization and multi-linear algebra methods in signal processing and data analysis, statistical and array signal processing, sparse signal processing, estimation, detection, and learning theory and methods, and multi-antenna, very large, cooperative, and cognitive systems. He was a member of the Sensor Array and Multi-Channel Signal Processing Committee of the IEEE Signal Processing Society from 2007 to 2012 and the Signal Processing for Communications and Networking Technical Committee of the IEEE Signal Processing Society from 2010 to 2016. He was a recipient of the 2004 IEEE Signal Processing Society Best Paper Award, the 2007 Alberta Ingenuity New Faculty Award, the 2011 Carl Zeiss Award (Germany), the 2012 NSERC Discovery Accelerator Award, and other awards. He was the Track Chair of Asilomar 2011, Pacific Grove, CA, USA; and the Tutorial Chair of ISWCS 2013, Ilmenau, Germany. He was the Technical Co-Chair of IEEE CAMSAP 2011, Puerto Rico; IEEE SAM 2018, Sheffield, U.K.; and IEEE CAMSAP 2021, Costa-Rica. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010 and IEEE SIGNAL PROCESSING LETTERS from 2007 to 2009. He has been a Senior Area Editor of IEEE SIGNAL PROCESSING LETTERS since 2016.

Publication III

A. K. Kocharalakota, S. A. Vorobyov and R. W. Heath. Attention Neural Network for Downlink Cell-Free Massive MIMO Power Control. In *Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp. 738-743, 2022.

© 2022

Reprinted with permission.

Attention Neural Network for Downlink Cell-Free Massive MIMO Power Control

1st Atchutaram K. Kocharalakota
Dept. Signal Processing and Acoustics
Aalto University
Espoo, Finland
kameswara.kocharalakota@aalto.fi

2nd Sergiy A. Vorobyov
Dept. Signal Processing and Acoustics
Aalto University
Espoo, Finland
svor@ieee.org

3rd Robert W. Heath Jr.
Dept. Electrical and Computer Engineering
North Carolina State University
Raleigh, North Carolina, USA
rwheathjr@ncsu.edu

Abstract—The downlink power control is challenging in a cell-free massive multiple-input multiple-output (CFmMIMO) system because of the non-convexity of the problem. This paper proposes a computationally efficient deep-learning algorithm to solve the max-min power control optimization problem subject to power constraints. To solve this problem, it presents an attention neural network (ANN) composed using the masked multi-head attention network modules, which are building blocks of the popular transformer neural network. The ANN solves the downlink power control problem of CFmMIMO in the presence of pilot contamination (non-orthogonal pilot sequences). The paper first translates the constrained optimization problem to an unconstrained one parameterized by the weights of the ANN. These weights are trained in an unsupervised fashion. The performance of the ANN power control algorithm is demonstrated using numerical simulations. The paper also provides a computational complexity analysis of the method.

I. INTRODUCTION

Cell-free massive multiple-input multiple-output (CFmMIMO) performs multiuser multiple-input multiple-output (MIMO) communications across distributed antennas or access points (APs). Decentralizing the antennas improves channel orthogonality and provides higher rates than colocated antennas, thanks to the differences in path loss between the users [1]–[7]. The time-division duplex (TDD) based CFmMIMO system with a large number of distributed antennas cooperating to serve a fewer number of users has been proposed in [8]. The CFmMIMO provides uniformly good service throughout the area of coverage. However, an efficient uplink/downlink power control algorithm is essential for achieving uniformly good service to all the users. Power control in CFmMIMO is challenging in general because of the computational complexity involved in solving the non-convex max-min fairness maximization problem [8], [9]. Specifically, designing a computationally efficient power control algorithm for the downlink signal at the central processing unit (CPU) is very challenging for CFmMIMO because of the number of optimization parameters involved in the downlink power control [9].

There are several methods for solving the power control problem using optimization or machine learning approaches. In order to achieve uniformly good spectral efficiency across the users, the power control algorithm proposed in [8] uses bisection search and a sequence of second-order cone program-

ming problems to find a solution. The algorithms in [10]–[13] use different second-order optimization techniques to solve the power control problem. Furthermore, a computationally efficient first-order optimization method called the accelerated proximal gradient method (APG) is proposed in [9].

Given the effectiveness of deep learning algorithms in computationally intense applications in communications [14]–[17], a study on the role of deep learning solutions for power control in CFmMIMO is essential. A fully connected neural network (FCN) for uplink power control of a CFmMIMO system is proposed in [18]–[20] while it is proposed for the more challenging case of downlink power control in [20]–[24]. Due to many optimization parameters involved, designing and training a scalable FCN for solving downlink power control is very challenging. Therefore, the deep-learning algorithms in the literature consider a CFmMIMO system with utmost 30 the number of APs and 12 number of users. Furthermore, these deep-learning algorithms made a simplistic assumption of having non-orthogonal pilots, causing no pilot contamination. Nevertheless, the authors in [22] propose an elegant convolutional neural network (CNN) algorithm for uplink power control in a CFmMIMO system that allows non-orthogonal pilots. Therefore, designing a computationally efficient deep-learning algorithm for downlink power control of a CFmMIMO system that allows non-orthogonal pilots is indispensable.

In this paper, we propose an attention neural network (ANN) composed using the masked multi-head attention network (MMHAN) modules, which are building blocks of the popular transformer neural network [25]. The ANN also takes pilot allocation information as input to compute the downlink power control coefficients along with the large-scale fading coefficients. We solve the max-min fairness maximization problem with power constraints by training the ANN in an unsupervised fashion. The power constraints are included in the optimization objective using a log-barrier approach. To the best of our knowledge, this is the first paper that solves a constrained downlink power control problem in a CFmMIMO system that allows non-orthogonal pilots. The motivation for the proposed algorithm is as follows.

- **Beyond FCN:** The power control problem considered in [18]–[21], [23], [24] accepts a 2-D fading coefficient

matrix as input to compute the optimal power control coefficients. The dimensions of this fading coefficient matrix are given by the number of APs \times the number of users. This fading coefficient matrix is flattened before feeding to an FCN, leading to a loss of the inherent structure in the input matrix. Correlation information between the columns of the fading coefficient matrix is particularly important in the case of pilot contamination. MMHAN modules in the ANN network proposed in this paper use the structure in this 2-D matrix by extracting the “attention” matrix (number of users \times the number of users). It captures the relation between the fading coefficient vectors corresponding to two different users (columns of the 2-D matrix).

- **Non-orthogonal pilots:** The masking feature in MMHAN also allows us to feed the mask input to the network. In this paper, we use this feature to feed the pilot allocation information to MMHAN modules. We calculate an interference matrix determined by all the users’ allocated pilot sequences and feed this matrix as the mask. Therefore, the final power control coefficients are computed based on the user channel correlations and pilot interferences. To the best of our knowledge, there is no deep-learning algorithm for the downlink power control problem that allows non-orthogonal pilots.
- **Scalable algorithm:** The hyper-parameters of ANN only depend upon the number of APs but not on the number of users. Therefore, ANN requires a fewer number of trainable parameters and smaller CPU memory than FCN. This allowed us to simulate the downlink power control for a medium-scale CFmMIMO network with 100 APs and 20 users.

II. SYSTEM MODEL

We consider a TDD CFmMIMO system where M APs serve K users in a large coverage area. Each user’s transceiver is equipped with a single antenna, and each AP’s transceiver is equipped with N antennas. We assume perfect synchronization and a narrowband channel.

The APs coordinate with each other through a fronthaul network and a CPU. The APs perform uplink channel estimation, uplink beamforming, and downlink beamforming operations. Along with the uplink symbol detection, the CPU performs slow rate baseband processing like pilot allocation, power control, and AP selection.

Each AP uses uplink training to estimate its uplink channel from all the users. The channel is assumed to be reciprocal so that the APs use the estimated uplink channel for uplink and downlink beamforming.

A. Channel Model

The channel from the k^{th} user to the m^{th} AP is modeled as $\mathbf{g}_{mk} = \beta_{mk}^{1/2} \mathbf{h}_{mk}$, where β_{mk} is the corresponding large-scale fading coefficient and $\mathbf{h}_{mk} \in \mathbb{C}^N$ is the vector of small-scale fading coefficients at each antenna of the AP (\mathbf{h}_{mk} is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I})$). The channel is assumed to be

constant within a coherence block (a block of time-frequency resources) of size τ symbols. The TDD CFmMIMO system performs uplink training, uplink data transmission, and downlink data transmission within every coherence block. Since this paper focuses on designing a computationally efficient downlink power control algorithm, we limit our system model description to uplink training and downlink data transmission. We refer the readers to [8] for a detailed description of the complete system model.

B. Uplink Training and Channel Estimation

Let τ_p ($\ll \tau$) be the number of symbols dedicated for the pilot transmission in each coherence block, and $\sqrt{\tau_p} \psi_k \in \mathbb{C}^{\tau_p}$ be the pilot sequence transmitted by the k^{th} user, such that $\|\psi_k\|_2 = 1$. Let ζ_p be the transmit signal to noise ratio (SNR) of each pilot symbol and $\mathbf{Z}_{p,m} \in \mathbb{C}^{N \times \tau_p}$ be the noise matrix whose elements are independently and identically distributed as $\mathcal{CN}(0, 1)$. Then, the uplink received signal at the m^{th} AP is given by

$$\mathbf{R}_m = \sqrt{\zeta_p \tau_p} \sum_{i=1}^K \mathbf{g}_{mi} \psi_i^T + \mathbf{Z}_{p,m}.$$

Assuming that the large-scale fading coefficients are known at the APs and the CPU, the MMSE estimate of the channel between the k^{th} user and the m^{th} AP can be straightforwardly shown to be

$$\hat{\mathbf{g}}_{mk} = \frac{\sqrt{\zeta_p \tau_p} \beta_{mk}}{1 + \zeta_p \tau_p \sum_{i=1}^K \beta_{mi} |\psi_i^H \psi_k|^2} \mathbf{R}_m \psi_k^* \quad (1)$$

and

$$\nu_{mk} = \mathbb{E}(|\hat{\mathbf{g}}_{mk}[n]|^2) = \frac{\zeta_p \tau_p \beta_{mk}^2}{1 + \zeta_p \tau_p \sum_{i=1}^K \beta_{mi} |\psi_i^H \psi_k|^2}$$

is the mean square value of the n^{th} element in $\hat{\mathbf{g}}_{mk}$. Note that ν_{mk} is the same for all the antenna elements $1 \leq n \leq N$.

C. Downlink Data Transmission

Here, we consider the matched filter beamforming. Let the downlink payload data symbol to the k^{th} user be c_k such that $\mathbb{E}\{c_k\} = 1$. Let ζ_d be the maximum transmit SNR of each data symbol normalized to the noise power, and μ_{mi} be the power control coefficient of the signal targeted to the i^{th} user from the m^{th} AP. Then, the beamformed downlink signal at the m^{th} AP is constructed as

$$\mathbf{x}_m = \sqrt{\zeta_d} \sum_{i=1}^K \frac{\mu_{mi}}{\sqrt{\nu_{mi}}} \hat{\mathbf{g}}_{mi}^* c_i.$$

Note that the total transmit power at the m^{th} AP is given by $\mathbb{E}\{\|\mathbf{x}_m\|_2^2\} = \zeta_d N \sum_{i=1}^K \mu_{mi}^2$.

Let $z_{d,m}$ be the noise signal distributed as $\mathcal{CN}(0, 1)$. Then, the received signal at the k^{th} user is given by

$$r_k = \sum_{m=1}^M \mathbf{g}_{mk}^T \mathbf{x}_m + z_{d,m}. \quad (2)$$

We now present the downlink spectral efficiency of a user in the CFmMIMO system. Let us define $\boldsymbol{\mu}_m \triangleq [\mu_{m1}, \dots, \mu_{mK}]$ as the vector of power control coefficients associated with the m^{th} AP, and the matrix of all power control coefficients as $\mathbf{M} \triangleq [\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_M^\top]^\top$. Note that the power control coefficients are positive real numbers that satisfy the downlink power constraint $\mathbb{E}\{\|\mathbf{x}_m\|_2^2\} \leq \zeta_d$, or equivalently, $\|\boldsymbol{\mu}_m\|_2^2 = \sum_{i=1}^K \mu_{mi}^2 \leq 1/N$, where $1 \leq m \leq M$. We define \mathbf{D}_k as a diagonal matrix whose m^{th} diagonal element is $\sqrt{\beta_{mk}}$, $\bar{\boldsymbol{\mu}}_i \triangleq [\mu_{1i}, \dots, \mu_{Mi}]^\top$ is the vector of power control coefficients associated with the i^{th} user, and

$$\boldsymbol{\nu}_{ik} \triangleq |\boldsymbol{\psi}_k^\top \boldsymbol{\psi}_i^*| \left[\sqrt{\nu_{1i}} \frac{\beta_{1k}}{\beta_{1i}}, \dots, \sqrt{\nu_{Mi}} \frac{\beta_{Mk}}{\beta_{Mi}} \right]^\top.$$

Using the well know use-and-then-forget capacity bounding technique in the mMIMO and CFmMIMO literature [8]–[10], [26], we can derive the following signal to interference and noise ratio (SINR), and a lower bound on downlink spectral efficiency (bits/s/Hz) of the k^{th} user¹.

$$\gamma_k(\mathbf{M}) = \frac{\zeta_d (\bar{\boldsymbol{\mu}}_k^\top \boldsymbol{\nu}_{kk})^2}{\zeta_d \sum_{i=1, i \neq k}^K (\bar{\boldsymbol{\mu}}_i^\top \boldsymbol{\nu}_{ik})^2 + \frac{\zeta_d}{N} \sum_{i=1}^K \|\mathbf{D}_k \bar{\boldsymbol{\mu}}_i\|_2^2 + \frac{1}{N^2}} \quad (3)$$

$$SE_k(\mathbf{M}) = \left(1 - \frac{\tau_p}{T}\right) \log_2(1 + \gamma_k(\mathbf{M})) \quad (4)$$

D. Problem Statement

We assume that the slowly varying large-scale fading coefficients are known at the APs and the CPU. It is shown in [8] that the CFmMIMO provides uniformly good service to all the users by using max-min power control. Therefore, we now formulate the max-min fairness maximization problem for the downlink power control subject to power constraints $\|\boldsymbol{\mu}_m\|_2^2 \leq \frac{1}{N}, 1 \leq m \leq M$. Let $S \triangleq \{\mathbf{M} | \mathbf{M} \geq \mathbf{0}; \|\boldsymbol{\mu}_m\|_2^2 \leq \frac{1}{N}, 1 \leq m \leq M\}$. Then, the constrained max-min fairness downlink power control problem can be formulated as

$$\begin{aligned} & \underset{\mathbf{M}}{\text{maximize}} && \min_{1 < k < K} SE_k(\mathbf{M}) \\ & \text{subject to} && \mathbf{M} \in S. \end{aligned} \quad (5)$$

Successive convex approximation (SCA) has been the standard technique for such types of quasi-convex problems. Such convex problems are solved using off-the-shelf second-order methods like interior point methods. In order to reduce the computational complexity involved in the second-order methods used in SCA solvers, the first-order APG method is proposed in [9] to solve the optimization problem efficiently. In the following section, we propose an ANN-based unsupervised learning method to further reduce the computational complexity in performing the downlink power control.

¹The downlink spectral efficiency for a similar setup is derived in multiple CFmMIMO papers. So, we omit the derivation and refer the readers to [9], [10].

III. ANN-BASED UNSUPERVISED LEARNING POWER CONTROL

The objective function of the constrained optimization problem (5) is non-differentiable. In this section, we first translate the constrained optimization problem to an unconstrained one with a differentiable objective function. Subsequently, we present the ANN algorithm.

A. Problem Reformulation

Let the large-scale fading coefficients matrix be

$$\mathbf{B} \triangleq \begin{bmatrix} \beta_{11} & \dots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{M1} & \dots & \beta_{MK} \end{bmatrix}. \quad (6)$$

Let Φ be the $K \times K$ interference matrix whose (i, k) element is given by $\Phi_{ik} \triangleq |\boldsymbol{\psi}_k^\top \boldsymbol{\psi}_i^*|^2$. Note that $\boldsymbol{\nu}_{ik}$ and hence SE_k are functions that depend on \mathbf{B} and Φ . Therefore, we represent the spectral efficiency function as $SE_k(\mathbf{B}, \Phi, \mathbf{M})$. The utility function that needs to be maximized can be defined as

$$\begin{aligned} U(\mathbf{B}, \Phi, \mathbf{M}, \mu_0) \triangleq & -\frac{1}{\lambda} \ln \left(\frac{1}{K} \sum_{k=1}^K \exp(-\lambda SE_k(\mathbf{B}, \Phi, \mathbf{M})) \right) \\ & + \frac{\eta}{2} \sum_{m=1}^M \ln \left(\frac{1}{N} - \|\boldsymbol{\mu}_m\|_2^2 - \mu_0^2 \right). \end{aligned} \quad (7)$$

Note that we employ the well-known log-barrier approach to ensure that the power constraints are met, and η is a regularization parameter that can be designed empirically. Furthermore, we convert the hard-minimum objective function in (5) to a soft-minimum function such that the optimization algorithm can acquire gradients, and λ is the smoothing parameter. We introduce an arbitrary optimization parameter μ_0 to avoid numerical issues (μ_0^2 can be seen as a slack variable). The corresponding unconstrained optimization problem can be written as

$$\begin{aligned} & \underset{\mathbf{M}, \mu_0}{\text{maximize}} && -\frac{1}{\lambda} \ln \left(\frac{1}{K} \sum_{k=1}^K \exp(-\lambda SE_k(\mathbf{B}, \Phi, \mathbf{M})) \right) \\ & && + \frac{\eta}{2} \sum_{m=1}^M \ln \left(\frac{1}{N} - \|\boldsymbol{\mu}_m\|_2^2 - \mu_0^2 \right). \end{aligned} \quad (8)$$

B. ANN formulation

1) *An Overview of MMHAN*: In their seminal work [25], the authors introduced the popular transformer neural network for performing natural language translation. They employ the attention mechanism to extract the coherence between the input vectors corresponding to the words in a sentence. The transformer neural network is composed using two vital building blocks: a multi-head attention-network module (MHAN) and an MMHAN module. These modules extract the coherence between the words (termed "self-attention") to accomplish the translation task. Compared to MHAN,

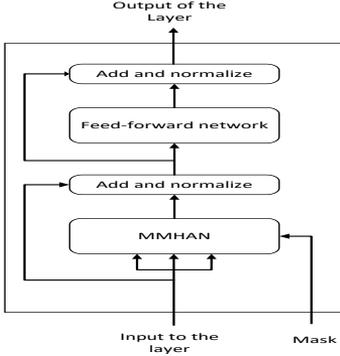


Fig. 1. Attention Layer

MMHAN performs additional masking functionality to ensure the causality of the translation work in that specific application. This paper uses the MMHAN module for building ANN to perform the downlink power control.

Here, we describe the mathematical representation of an MMHAN module. As the name suggests, MMHAN performs the attention mechanism using multiple heads of identical operations. Let the number of heads be H such that M is an integer multiple of H ($M = HD$). Let $\mathbf{X} \in \mathbb{R}^{K \times M}$ be the input matrix to MMHAN and $\mathbf{Y} \in \mathbb{R}^{K \times M}$ be the output. The MMHAN module also takes a $K \times K$ mask matrix as input. In this paper, we consider Φ as the mask matrix. The input matrix \mathbf{X} is split into H matrices and fed as input to each head of MMHAN. The outputs of each head are concatenated to form the output matrix \mathbf{Y} . The input and output matrices of the head h ($1 \leq h \leq H$) are given by

$$\begin{aligned} \mathbf{X}_{\{k,m\}}^{(h)} &= \mathbf{X}_{\{k,(m+(h-1)D)\}} & 1 \leq k \leq K, 1 \leq m \leq D \\ \mathbf{Y}_{\{k,m\}}^{(h)} &= \mathbf{Y}_{\{k,(m+(h-1)D)\}} & 1 \leq k \leq K, 1 \leq m \leq D. \end{aligned}$$

Then, the output of the head h is computed as follows

$$\mathbf{Y}^{(h)} = \text{softmax}(\Phi \odot \mathbf{A})\mathbf{V} \quad (9)$$

where $\mathbf{A} = \mathbf{Q}\mathbf{K}^T/\sqrt{D}$, $\mathbf{Q} = \mathbf{X}^{(h)}\mathbf{W}_Q^{(h)}$, $\mathbf{K} = \mathbf{X}^{(h)}\mathbf{W}_K^{(h)}$, and $\mathbf{V} = \mathbf{X}^{(h)}\mathbf{W}_V^{(h)}$. Here, \odot represents the elementwise matrix multiplication. Note that $\mathbf{W}_Q^{(h)}$, $\mathbf{W}_K^{(h)}$, and $\mathbf{W}_V^{(h)}$ are trainable parameters (matrices of dimension $D \times D$). Finally, we represent the mathematical relation between the input and output matrices of the MMHAN as the following

$$\mathbf{Y} = f_M(\mathbf{X}, \Phi, \mathbf{W}_M)$$

where \mathbf{W}_M is the set of all trainable matrices corresponding to all the heads of MMHAN.

2) *ANN structure*: Using the MMHAN module described in the above subsection, we first define an attention layer which is used for composing ANN. Fig. 1 depicts the structure

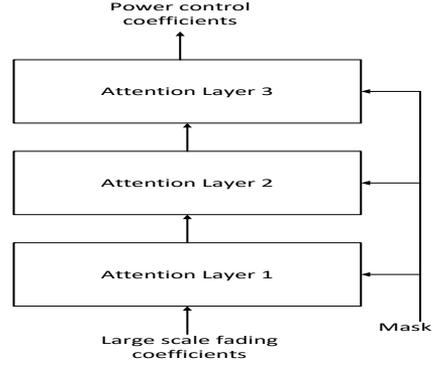


Fig. 2. ANN structure.

of the attention layer. Let \mathbf{X} be the input and \mathbf{Z} be the output of an attention layer. The relation between input and output of the layer can be mathematically represented using following

$$\begin{aligned} \mathbf{Y} &= f_M(\mathbf{X}, \Phi, \mathbf{W}_M) \\ \tilde{\mathbf{Y}} &= \text{LayerNorm1}(\mathbf{X} + \mathbf{Y}, \mathbf{W}_1) \\ \hat{\mathbf{Y}} &= \text{FF}(\tilde{\mathbf{Y}}, \mathbf{W}_2) \\ \mathbf{Z} &= \text{LayerNorm2}(\tilde{\mathbf{Y}} + \hat{\mathbf{Y}}, \mathbf{W}_3) \end{aligned}$$

where, the FF is a feed forward neural network operates with one hidden layer with $D_m \text{in} = 4M$ number of hidden units with rectified linear unit activation. \mathbf{W}_2 represent the weights of the feed-forward network. This feed-forward network operates row-wise, i.e., the feed forward operation of the same network is applied on all the rows. Note that the attention layer uses two layer-normalization layers with trainable parameters \mathbf{W}_1 and \mathbf{W}_3 . $\tilde{\mathbf{Y}}$ and $\hat{\mathbf{Y}}$ are intermediate outputs.

Now, we propose an ANN network composed of 3 attention layers taking the same mask matrix given by Φ as input. Fig. 2 depicts the ANN architecture.

C. ANN for the Downlink Power Control

Let the preprocessed input matrix be $\tilde{\mathbf{B}} \triangleq \ln \mathbf{B}^T$, where the logarithm operation is performed element-wise. Consider an ANN algorithm that takes $\tilde{\mathbf{B}}$ as input and outputs the power control coefficient matrix \mathbf{M} .

Let us model the power control coefficient matrix as a function of $\tilde{\mathbf{B}}$ parameterized by the weights \mathbf{W} as

$$\mathbf{M} = f_{ANN}(\tilde{\mathbf{B}}, \Phi, \mathbf{W}). \quad (10)$$

D. Unsupervised Learning

A data-driven approach for efficiently solving the power control problem for a CFmMIMO system is the following. Assuming that $\tilde{\mathbf{B}}$ is a random matrix that follows an unknown distribution, in this approach, the optimization problem can be reformulated as

$$\underset{\mathbf{W}, \mu_0}{\text{maximize}} \quad \mathbb{E}_{\tilde{\mathbf{B}}} \left[U \left(\tilde{\mathbf{B}}, \Phi, f(\tilde{\mathbf{B}}, \Phi, \mathbf{W}), \mu_0 \right) \right]. \quad (11)$$

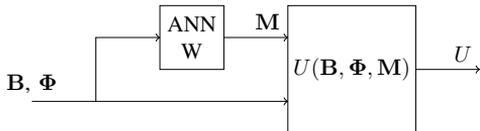


Fig. 3. ANN base power control algorithm.

Assume that the AP placement and the path loss model are given. Then, to solve the optimization problem, for a given AP placement and path loss model, we generate a large set of P fading coefficient matrices given by

$$\{\mathbf{B}^{(p)} \in \mathbb{R}_+^{M \times K} | 1 \leq p \leq P\}$$

corresponding to P uniform random placements of the K users. Furthermore, we consider that each user picks a random pilot sequence from a set of K orthogonal pilots. Using an off-the-shelf ADAM optimizer, we train the ANN (learn the parameters \mathbf{W}) by maximizing the empirical average of the utility function. According to the law of large numbers, the empirical average approaches the ensemble average of the utility function and solves the problem in (11). Fig. 3 depicts the ANN-based power control algorithm setup used in this paper.

Although training of this network involves a high computational cost, the computational cost in the deployment phase is $\mathcal{O}(M^2K + HD^2K + HDK^2)$ since it involves only one forward pass of the ANN. This is a significant saving compared to state-of-the-art algorithms. For example, the most computationally efficient state-of-the-art optimization-based method, that is, the APG method [9], solves the problem with the computational cost of $\mathcal{O}(MK^2)$ per iteration, and the number of iterations for such type of first-order methods is thousands to tens or even hundreds of thousands. Therefore, with appropriate neural network training, the computational complexity of a power control algorithm can be significantly reduced compared to conventional optimization algorithms.

IV. SIMULATIONS

Consider a CFmMIMO system with $M = 100$ APs and $K = 20$ users. Each AP consists of a single antenna ($N = 1$). All the APs are spread over an area of $D \times D$ km² where $D = 1$. The lengths of the coherence interval and of the pilot training are $\tau = 200$ symbols and $\tau_p = 20$ symbols, respectively.

Assume for the noise that $N_f = 9$ dB and the power spectral density $N_0 = -173.98$ dBm/Hz. The bandwidth of the channel is $BW = 20$ MHz. Then, the total noise power is given by $P_n = BW10^{(N_0 + N_f - 30)/10}$ W. Therefore, we consider the transmit SNR of the uplink pilot and the downlink data as $\zeta_p = 0.2/P_n$ and $\zeta_d = 1/P_n$, respectively.

For the entire simulations, the AP locations are fixed while P samples of \mathbf{B} 's are generated by varying the user locations. Let d_{mk} be the distance between the m^{th} AP and the k^{th} user in [km]. Consider also a three-slope path loss model, in which $PL_{mk} = -L - 15 \log_{10}(d_1) - 20 \log_{10}(d'_{mk})$ is the path loss

(in [dB]) of the channel between the m^{th} AP and the k^{th} user, where

$$d'_{mk} = \begin{cases} d_0 & d_{mk} \leq d_0 \\ d_{mk} & d_0 < d_{mk} \leq d_1 \\ d_1 & d_{mk} > d_1. \end{cases} \quad (12)$$

Finally, we model the large-scale fading coefficient (in [dB]) of the channel between the m^{th} AP and the k^{th} user as $\beta_{mk} = PL_{mk} + z_{mk}$, where z_{mk} represents shadow fading that follows the normal distribution with mean 0 and variance σ_{sh}^2 . Using similar setup as in [8], choose $L = 140.72$ dB, $d_0 = 0.01$ km, $d_1 = 0.05$ km, and $\sigma_{\text{sh}} = 8$ dB.

The ANN is trained using $P = 32000$ samples of \mathbf{B} . Let the learning rate be equal 10^{-4} . The batch size is 16. In our simulations, choose $\eta = 10^{-5}$ and $\lambda = 3$. After training the ANN, a set of 200 samples of \mathbf{B} are used to test the performance of the trained network. Fig. 4 shows the comparison of the performance of the trained ANN with that of the simple equal power allocation (EPA) algorithm, FCN, and the APG algorithm. The figure plots the empirical cumulative distribution function (CDF) of per-user spectral efficiency in the network. Furthermore, Table I shows the average computational time required for each algorithm. The ANN-based algorithm clearly outperforms the EPA algorithm. Note that the FCN does not meet the performance of the EPA algorithm. This is due to the introduction of non-orthogonal pilots, and FCN is ignorant of this information. Note that the proposed ANN algorithm approaches the performance of the APG with a great reduction in computational time (with proper fine-tuning, we might reach the performance of APG. We plan this for the final draft). Therefore, we can conclude that ANN significantly reduces the computational complexity of power allocation at the cost of a slight degradation in the average spectral efficiency of the users in the system.

V. CONCLUSION

The problem of allocating optimal downlink power control coefficients for a CFmMIMO system has been formulated as an unconstrained optimization problem using a log-barrier approach. An ANN network is proposed for solving the max-min fairness optimization problem by training it in an unsupervised fashion such that the expected utility function is maximized. Using simulation results, we have demonstrated that the ANN algorithm approaches the state-of-the-art APG power control algorithm's performance and significantly reduces computational complexity. This is achieved by extracting correlation in large-scale fading coefficient vectors of different users and also by providing a way to deal with non-orthogonal pilots.

| Algorithm | Run-time (in secs) |
|-----------|--------------------|
| EPA | 0.000074705 |
| APG | 2.875497515 |
| FCN | 0.001639365 |
| ANN | 0.003220985 |

TABLE I
RUN-TIME

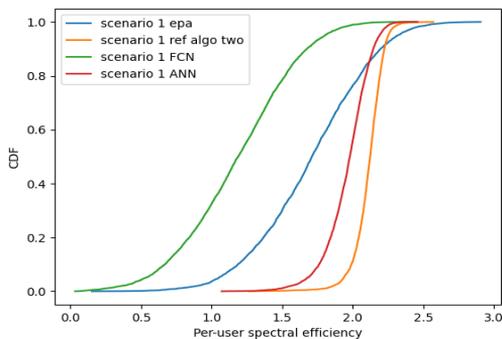


Fig. 4. CDFs of per user spectral efficiency in CFmMIMO for different power control algorithms.

REFERENCES

- [1] K. T. Truong and R. W. Heath, "The viability of distributed antennas for massive mimo systems," in *Proc. Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2013*, pp. 1318–1323.
- [2] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 161–167, 2013.
- [3] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, 2006.
- [4] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 69–73, 2007.
- [5] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.
- [6] X. Hong, Y. Jie, C.-X. Wang, J. Shi, and X. Ge, "Energy-spectral efficiency trade-off in virtual mimo cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2128–2140, 2013.
- [7] D. Castanheira and A. Gameiro, "Distributed antenna system capacity scaling [coordinated and distributed mimo]," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 68–75, 2010.
- [8] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive mimo versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.
- [9] M. Farooq, H. Q. Ngo, and L. N. Tran, "Accelerated projected gradient method for the optimization of cell-free massive mimo downlink," in *Proc. IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK, 2020*, pp. 1–6.
- [10] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive mimo," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 25–39, 2018.
- [11] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive mimo with zero-forcing precoding design," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1871–1874, 2017.
- [12] S. Buzzi and A. Zappone, "Downlink power control in user-centric and cell-free massive mimo wireless networks," in *Proc. IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, Montreal, QC, Canada, 2017*, pp. 1–6.
- [13] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive mimo: A blessing in disguise," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5153–5169, 2019.
- [14] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2554–2564, 2019.
- [15] R. M. Dreifuerst, R. W. Heath, M. N. Kulkarni, and J. Charlie, "Deep learning-based carrier frequency offset estimation with one-bit adcs," in *Proc. IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications, Atlanta, GA, USA, 2020*, pp. 1–5.
- [16] E. Balevi, A. Doshi, and J. G. Andrews, "Massive mimo channel estimation with an untrained deep neural network," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2079–2090, 2020.
- [17] Y. Zhang, A. Doshi, R. Liston, W.-T. Tan, X. Zhu, J. G. Andrews, and R. W. Heath, "Deepwiphy: Deep learning-based receiver design and dataset for ieee 802.11ax systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1596–1611, 2021.
- [18] N. Rajapaksha, K. B. Shashika Manosha, N. Rajatheva, and M. Latva-Aho, "Deep learning-based power control for cell-free massive mimo networks," in *Proc. IEEE International Conference on Communications, Montreal, QC, Canada, 2021*, pp. 1–7.
- [19] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised-learning power control for cell-free wireless systems," in *Proc. IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey, 2019*, pp. 1–5.
- [20] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised learning for c-ran power control and power allocation," *IEEE Communications Letters*, vol. 25, no. 3, pp. 687–691, 2021.
- [21] S. Chakraborty, E. Björnson, and L. Sanguinetti, "Centralized and distributed power allocation for max-min fairness in cell-free massive mimo," in *Proc. Asilomar Conference on Signals, Systems, and Computers, 2019*, pp. 576–580.
- [22] M. Bashar, A. Akbari, K. Cumanan, H. Q. Ngo, A. G. Burr, P. Xiao, M. Debbah, and J. Kittler, "Exploiting deep learning in limited-fronthaul cell-free massive mimo uplink," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1678–1697, 2020.
- [23] Y. Zhao, I. G. Niemegeers, and S. H. De Groot, "Power allocation in cell-free massive mimo: A deep learning method," *IEEE Access*, vol. 8, pp. 87 185–87 200, 2020.
- [24] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised-learning power allocation for the cell-free downlink," in *Proc. IEEE International Conference on Communications, Dublin, Ireland, 2020*, pp. 1–5.
- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.
- [26] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge: Cambridge University Press, 2016.

Publication IV

A. K. Kocharlakota, S. A. Vorobyov and R. W. Heath. Pilot Contamination Aware Transformer Neural Network for Downlink Power Control in Cell-Free Massive MIMO Networks. *Submitted to IEEE Trans. Wireless Commun.*, Jun 2024.

© 2024

Reprinted with permission.

Pilot Contamination Aware Transformer Neural Network for Downlink Power Control in Cell-Free Massive MIMO Networks

Atchutaram K. Kocharalakota, *Student Member, IEEE*, Sergiy A. Vorobyov, *Fellow, IEEE*,
and Robert W. Heath Jr., *Fellow, IEEE*

Abstract—Learning-based downlink power control in cell-free massive multiple-input multiple-output (CFmMIMO) systems offers a promising alternative to conventional iterative optimization algorithms, which are computationally intensive due to online iterative steps. However, existing learning-based methods often fail to exploit the intrinsic structure of channel data and neglect pilot allocation information, leading to suboptimal performance, especially in large-scale networks with many users. This paper introduces the pilot contamination-aware power control (PAPC) transformer neural network, a novel approach that integrates pilot allocation data into the network, effectively handling pilot contamination scenarios. PAPC employs the attention mechanism with a custom masking technique to utilize structural information and pilot data. The architecture includes tailored preprocessing and post-processing stages for efficient feature extraction and adherence to power constraints. Trained in an unsupervised learning framework, PAPC is evaluated against the accelerated proximal gradient (APG) algorithm, showing comparable spectral efficiency fairness performance while significantly improving computational efficiency. Simulations demonstrate PAPC’s superior performance over fully connected networks (FCNs) that lack pilot information, its scalability to large-scale CFmMIMO networks, and its computational efficiency improvement over APG. Additionally, by employing padding techniques, PAPC adapts to the dynamically varying number of users without retraining.

Index Terms—Large-Scale Cell-Free Massive MIMO (CFmMIMO), Pilot Contamination, Transformer Neural Network, Pilot Contamination-Aware Power Control (PAPC), Generative Pretrained Transformer (GPT), Bidirectional encoder Representations from Transformers (BERT), Downlink Power Control, Deep Learning.

I. INTRODUCTION

Base station (BS) coordination eliminates inter-cell interference and allows multi-user massive multiple-input multiple-output (MIMO) to serve users distributed over a large geographic area. Such coordination has been explored to increase per-user data rates and spectral efficiency (SE) of systems that are referred in the literature by different (not equivalent but having some uncommon features) terms such as distributed

Initial part of this work was presented at 56th Asilomar Conference on Signals Systems, and Computers, Asilomar, CA, USA, Nov. 2022.

A. K. Kocharalakota and S. A. Vorobyov are with the Department of Information and Communications Engineering, Aalto University, PO Box 15400, 00076 Aalto, Finland. (e-mails: kameswara.kocharalakota@aalto.fi and sergiy.vorobyov@aalto.fi). A. K. Kocharalakota is also with Nokia, Espoo, Finland.

R. W. Heath Jr. is with the Department of Electrical and Computer Engineering, University of California San Diego, 9500 Gillman Dr, CA, USA 92093-0021. (e-mail: rwheathjr@ucsd.edu).

multi-user MIMO antenna systems [1, 2], cloud radio access networks (cloud RANs) [3, 4], and cell-free massive MIMO (CFmMIMO) systems [5–8] as used in this paper.

To fully leverage the benefits of BS coordination, sophisticated pilot allocation and power control algorithms are essential. These algorithms face significant computational complexities due to the centralized signal processing tasks required at a central processor [9–11]. Specifically, designing a downlink power control algorithm involves a large number of optimization parameters, posing a significant obstacle in the development of CFmMIMO infrastructure [5, 12]. Addressing non-convex optimization problems is the core challenge.

Power control algorithms in CFmMIMO typically employ various objective functions and second-order interior-point methods for optimization [5, 13–15]. For instance, the max-min fairness optimization was tackled using bisection search methods and second-order cone feasibility problems in [5]. Similarly, energy efficiency maximization was addressed through successive convex approximation techniques in [13, 14], which were also applied to the max-min fairness problem in [15].

The scalability of second-order interior-point methods for power control in CFmMIMO systems remains a key challenge [12, 15–18]. Two strategies have been proposed to address this issue: user-centric approaches [15–17] and first-order power control algorithms [12, 18], both offering reduced complexity. The user-centric approach enhances efficiency by grouping BSs and setting large-scale fading coefficients between users and non-group BSs to zero, thereby reducing the computational burden. On the other hand, first-order methods reduce computational complexity by employing efficient solvers while operating on network-wide large-scale fading information without BS grouping, similar to second-order methods.

A. Related literature

The shift from second-order to first-order methods like accelerated projected gradient (APG) [12] aimed to improve scalability, but computational challenges persisted due to reliance on online iterative solvers, especially in large-scale networks. User-centric approaches also employed similar online solvers.

To overcome these challenges, learning-based solutions were introduced. Trained offline to reduce complexity during

inference, these methods map the input large-scale fading coefficients directly to output power control coefficients. This eliminates online iterative solvers, reducing computational complexity [19–38].

Various learning-based solutions were proposed for uplink and downlink power control in CFmMIMO systems [19–35]. A fully connected network (FCN) in an unsupervised learning setup was introduced in [19–21] for uplink power control with different objectives. A long short-term memory network was proposed in [22] under supervised learning, while FCNs were explored in [23, 24]. Reinforcement learning (RL) variants were proposed in [25–28]. For downlink power control, unsupervised FCN solutions were studied in [29, 30], and an RL solution was explored in [31]. Both uplink and downlink were addressed by unsupervised FCN in [32]. Supervised convolutional neural network (CNN)-based solutions were proposed in [33, 34], and a graph neural network (GNN)-based solution for downlink appeared in [35]. These algorithms commonly assumed orthogonal pilots and contamination-free environments.

The studies in [36–38] proposed deep learning solutions that considered pilot contamination scenarios during testing. A distributed unsupervised FCN for downlink power control was introduced in [36], while a similar solution for both uplink and downlink was presented in [37]. In [38], a distributed RL solution for downlink power control was proposed. Although these models were tested with pilot contamination, they did not fully address pilot contamination during training or model design.

The effectiveness of neural network architectures for downlink power control depends on how they handle network-wide large-scale fading coefficients. FCN methods flatten the large-scale fading coefficient matrix (number of BSs \times number of users), losing key associations between BSs and users. While CNN-based methods [33, 34] preserve the matrix structure, their use is debatable since these matrices lack the localized clustering seen in natural images. In contrast, the GNN architecture in [35] effectively leverages the structural information, providing a better solution for power control.

Pilot contamination remains a key challenge despite advancements in learning-based downlink power control methods. Ignoring pilot contamination limits the power control algorithms to small-scale CFmMIMO systems, restricting the number of users and impeding scalability due to the limited number of available orthogonal pilots. None of the methods discussed, including the GNN, directly address pilot contamination, often assuming ideal conditions. Unlike traditional optimization techniques, learning-based solutions lack explicit handling of pilot contamination, revealing a gap in addressing this issue while preserving the structure in the channel.

B. Contributions

This paper introduces a novel Pilot Contamination-Aware Power Control (PAPC) transformer neural network designed for downlink power control in a CFmMIMO network. PAPC leverages the large-scale fading coefficient matrix and a newly formulated matrix representation of pilot allocation information to map these inputs directly to power control coefficients.

It is trained in an unsupervised fashion to maximize the empirical smoothed-minimum per-user spectral efficiency under power constraints, using modified transformer blocks with additional preprocessing and postprocessing modules. During inference, the model addresses the max-min fairness downlink power control problem. Training and testing are conducted in a time-division duplexed (TDD) CFmMIMO system with minimum mean square error channel estimation (MMSE) and matched filter downlink beamforming.¹ The contributions of this paper are summarized as follows.

- **Attention mechanism for exploiting inter-user relationships:** The proposed PAPC transformer preserves the structural integrity of the network-wide two-dimensional large-scale fading coefficient matrix, avoiding the common pitfall of the flattening operation that can destroy crucial inter-user relationships. The attention mechanism in the transformer blocks of PAPC plays a vital role in extracting these relationships among the users. By doing so, it effectively learns to handle the dynamics of the propagation environment.
- **Modification of transformer blocks to incorporate pilot contamination matrix:** Without utilizing prior information from the pilot allocation algorithm, the power control algorithm’s effectiveness is compromised. To address this, the standard transformer blocks are modified with a novel masking mechanism that incorporates the pilot contamination matrix into the neural network. This integration allows the PAPC to efficiently handle pilot contamination scenarios and provide enhanced inference accuracy in both contaminated and uncontaminated environments.
- **Enhancing accuracy with preprocessing and postprocessing stages:** The PAPC model’s accuracy is enhanced by adding preprocessing and postprocessing stages. Preprocessing increases input dimensionality, allowing the model to learn richer features from each user’s large-scale fading coefficients. Postprocessing adjusts the output power control coefficients back to the desired dimensionality while ensuring the compliance to power constraints.
- **Enabling adaptability:** The customized architecture of the postprocessing module, combined with padding techniques, allows the PAPC to handle varying numbers of users. This customization enables the model to adapt to different sizes of CFmMIMO without requiring a redesign or retraining.
- **Scalability through reduced hyperparameters:** The PAPC transformer’s hyperparameters depend only on the number of BSs, unlike the FCN model, which also depends on the number of users. This reduction in hyperparameters enhances the scalability and efficiency in large CFmMIMO networks. The model’s pilot contamination awareness also makes it suitable for handling a large number of users in these networks. The scalability of the PAPC model has been validated through extensive testing

¹The model should be adaptable to other techniques and optimization objectives, though this is not directly tested in this work as it is outside of the scope.

in large-scale CFmMIMO settings with up to 100 APs and 80 users.²

- **Benchmarking against advanced algorithms:** The performance of the PAPC model is validated against the network-wide first-order APG method from [12]. Known for its computational efficiency, the APG method converged about 100 times faster in large-scale CFmMIMO systems, similar to those in this paper, compared to second-order methods. It serves as a strong benchmark for assessing the PAPC model. By targeting comparable performance to APG while significantly reducing computational complexity, the PAPC model demonstrates its potential for enabling large-scale CFmMIMO deployments.

Compared with prior work on centralized and distributed deep neural network (DNN)-based downlink power control algorithms [29, 30, 32–37], PAPC not only leverages inter-user relationships within the network-wide large-scale fading coefficient matrix in a centralized manner but also accounts for pilot contamination scenarios by incorporating pilot allocation information. To the best of our knowledge, this is the first work to propose DNN-based pilot contamination-aware downlink power control algorithms in CFmMIMO.

An earlier development in this direction was presented in a conference publication [39], but this paper significantly improves upon it by enhancing accuracy through improved feature extraction, introducing new preprocessing and post-processing stages, and improved hyperparameter tuning.

C. Paper Organization

The paper is organized as follows. Section II introduces the system model and provides the downlink SE expression for CFmMIMO networks. Section III highlights challenges in designing a DNN for downlink power control, particularly addressing pilot contamination. Section IV describes the proposed PAPC transformer architecture. Section V presents numerical evaluations and comparisons, and Section VI concludes the work with final observations.

D. Notation

The sets of real, positive real, and complex numbers are denoted by \mathbb{R} , \mathbb{R}_+ , and \mathbb{C} , respectively. Matrices are represented using boldface capital letters, while vectors are indicated by boldface lowercase letters. For matrix and vector operations, the superscripts $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ signify element-wise conjugate, transpose, and Hermitian transpose operations. Additionally, the symbol \odot represents the Hadamard product, an element-wise product of two matrices or vectors. Element-specific operations are denoted as a_i and $a_{i,j}$, representing the i^{th} element of vector \mathbf{a} and $(i,j)^{\text{th}}$ element of matrix \mathbf{A} , respectively, and $\mathbf{1}_K$ represents a K dimensional vector of ones. The function $\ln(\cdot)$ is used to denote the natural logarithm operation. For statistical notation, $\mathcal{N}_{\mathbb{C}}(m, \sigma^2)$ describes a circularly symmetric complex Gaussian random variable with mean vector m and variance σ^2 . The norm $\|\cdot\|$ indicates

²We follow the characterization of large-scale CFmMIMO systems from [37], where a system is categorized as large-scale if $MK \geq 1000$.

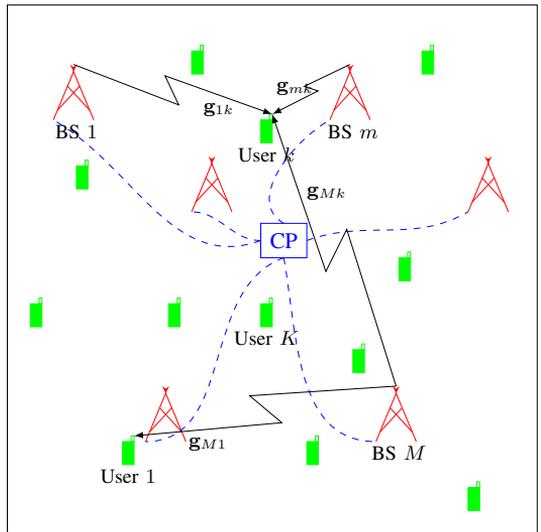


Fig. 1: Illustration of the CFmMIMO system with M distributed BSs, each equipped with N antennas, serving K single-antenna users under the coordination of a CP. The channel between BS m and user k is denoted by \mathbf{g}_{mk} .

the l_2 norm of a vector, and $\mathbb{E}[\cdot]$ denotes the mathematical expectation of a random variable.

II. SYSTEM MODEL

Consider a fully synchronized time-division duplexed (TDD) CFmMIMO system where M BSs, each with N antennas, serve K single-antenna users ($M > K$). All the BSs are connected to a central processor via a backhaul network and they simultaneously serve all users using common time-frequency resources, as depicted in Fig. 1.

Assuming block-fading channels that remain constant over a coherence block of τ symbols, the uplink channel from the k^{th} user to the m^{th} BS is $\mathbf{g}_{mk} = \sqrt{\beta_{mk}}\mathbf{h}_{mk}$, where β_{mk} is the large-scale fading coefficient and $\mathbf{h}_{mk} \in \mathbb{C}^N$ is the vector of small-scale fading coefficients with its elements independent and identically distributed (i.i.d.) as $\mathcal{N}_{\mathbb{C}}(0, 1)$. The large-scale fading coefficients β_{mk} , $\forall m, k$ are constant over many coherence blocks, while the small-scale fading coefficients \mathbf{h}_{mk} , $\forall m, k$ change each block. The large-scale fading coefficients are assumed to be known at all the BS's and the central processor.

During the uplink training phase, users transmit pilot sequences of length τ_p ($\tau_p \ll \tau$), where the k^{th} user's pilot is $\sqrt{\tau_p}\boldsymbol{\psi}_k \in \mathbb{C}^{\tau_p}$ with $\|\boldsymbol{\psi}_k\| = 1$. For $K > \tau_p$, due to limited number of orthogonal pilots, the sequences are reused, causing pilot contamination.

The transmit signal-to-noise ratio (SNR) per pilot symbol is ζ_p , and the additive white Gaussian noise (AWGN) at the m^{th} BS is $\mathbf{Z}_{p,m} \in \mathbb{C}^{N \times \tau_p}$ with i.i.d. elements following the

distribution $\mathcal{N}_C(0, 1)$. The received signal at the m^{th} BS is

$$\mathbf{R}_m = \sqrt{\zeta_p \tau_p} \sum_{i=1}^K \mathbf{g}_{mi} \psi_i^T + \mathbf{Z}_{p,m}.$$

It is straightforward to find that the MMSE estimate of the channel between the k^{th} user and the m^{th} BS is

$$\hat{\mathbf{g}}_{mk} = \frac{\sqrt{\zeta_p \tau_p} \beta_{mk}}{1 + \zeta_p \tau_p \sum_{i=1}^K \beta_{mi} |\psi_i^H \psi_k|^2} \mathbf{R}_m \psi_k^*,$$

and the mean square value of each element in $\hat{\mathbf{g}}_{mk}$ is

$$\nu_{mk} = \mathbb{E} [|\hat{\mathbf{g}}_{mk}[n]|^2] = \frac{\zeta_p \tau_p \beta_{mk}^2}{1 + \zeta_p \tau_p \sum_{i=1}^K \beta_{mi} |\psi_i^H \psi_k|^2},$$

which is uniform across all antenna elements n .

In the downlink data transmission phase, let the k^{th} user's data symbol be c_k with $\mathbb{E} [|c_k|^2] = 1$. Let ζ_d be the maximum transmit SNR per data symbol, and μ_{mi} be the power control coefficient for the signal to the i^{th} user from the m^{th} BS. The downlink transmit signal at the m^{th} BS is

$$\mathbf{x}_m = \sqrt{\zeta_d} \sum_{i=1}^K \frac{\mu_{mi}}{\sqrt{\nu_{mi}}} \hat{\mathbf{g}}_{mi}^* c_i.$$

Then, $\mathbb{E} [|\|\mathbf{x}_m\|^2|] = \zeta_d N \sum_{i=1}^K \mu_{mi}^2$ represents the transmit power at the m^{th} BS. The downlink received signal at the k^{th} user is

$$r_k = \sum_{m=1}^M \mathbf{g}_{mk}^T \mathbf{x}_m + z_{d,k}.$$

where $z_{d,k} \sim \mathcal{N}_C(0, 1)$ is the AWGN noise.

Define a vector of power control coefficients for the k^{th} user as $\bar{\boldsymbol{\mu}}_k = [\mu_{1k}, \dots, \mu_{Mk}]^T$, a diagonal matrix \mathbf{D}_k as $\mathbf{D}_k = \text{diag}(\sqrt{\beta_{1k}}, \dots, \sqrt{\beta_{Mk}})$, and a vector $\boldsymbol{\nu}_{ik} \in \mathbb{R}^M$ such that

$$\boldsymbol{\nu}_{ik}[m] = |\psi_k^T \psi_i^*| \sqrt{\frac{\beta_{mk}}{\beta_{mi}}}.$$

Using the use-and-then-forget bounding technique, the signal-to-interference-plus-noise ratio (SINR) for the k^{th} user is [12, 14]

$$\gamma_k = \frac{\zeta_d (\bar{\boldsymbol{\mu}}_k^T \boldsymbol{\nu}_{kk})^2}{\sum_{\substack{i=1 \\ i \neq k}}^K \zeta_d (\bar{\boldsymbol{\mu}}_i^T \boldsymbol{\nu}_{ik})^2 + \frac{\zeta_d}{N} \sum_{i=1}^K \|\mathbf{D}_k \bar{\boldsymbol{\mu}}_i\|^2 + \frac{1}{N^2}}, \quad (1)$$

and a lower bound on the downlink SE (bits/s/Hz) is

$$\text{SE}_k = \left(1 - \frac{\tau_p}{\tau}\right) \log_2(1 + \gamma_k). \quad (2)$$

Define the large-scale fading coefficients matrix as $\mathbf{B} \in \mathbb{R}_+^{M \times K}$. The element at position (m, k) of \mathbf{B} is β_{mk} . Let $\boldsymbol{\mu}_m = [\mu_{m1}, \dots, \mu_{mK}]^T$ denote the power control coefficients at the m^{th} BS. Define the matrix of all power control coefficients as

$$\mathbf{M} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1K} \\ \vdots & \ddots & \vdots \\ \mu_{M1} & \cdots & \mu_{MK} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \vdots \\ \boldsymbol{\mu}_M^T \end{bmatrix} = [\bar{\boldsymbol{\mu}}_1 \quad \cdots \quad \bar{\boldsymbol{\mu}}_K]. \quad (3)$$

Then the SE is a function of power control coefficients and pilots, while the large-scale fading coefficients are considered fixed parameters that characterize the function. Given the large-scale fading coefficients (matrix \mathbf{B}), the central processor is responsible to perform optimal pilot allocation (compute K pilots, ψ 's) and power control (compute matrix \mathbf{M}).

III. CHALLENGES IN DNN-BASED POWER CONTROL

A. Need for Learning-based Approaches

Max-min fairness optimization approaches in downlink power control aim to maximize the minimum SE across users. These methods target maximizing the SE of the worst-performing user to ensure fairness, while managing power distribution and pilot allocation across the network. The APG algorithm [5] is the state-of-the-art solution for power control in large-scale CFmMIMO systems, efficiently handling power control through a first-order accelerated method. Power control is subject to constraints that limit the total transmitted power per BS, expressed using the set:

$$\mathcal{S} = \left\{ \mathbf{M} \left| \begin{array}{l} \mathbf{M} \geq \mathbf{0}, \\ \|\boldsymbol{\mu}_m\|^2 \leq \frac{1}{N}, \\ \forall m \in \{1, \dots, M\} \end{array} \right. \right\}. \quad (4)$$

While traditional methods rely on iterative solvers, learning-based approaches offer an alternative to mitigate the computational costs [19–38]. In supervised learning, large datasets are used to train DNN models that approximate the unknown function mapping inputs of the iterative algorithms to the output power control coefficients. Despite lacking a closed-form expression, learning-based methods can efficiently learn this mapping through extensive training.

Unsupervised learning, unlike supervised methods, does not rely on reference solutions generated by iterative algorithms. Both approaches perform function approximation, but in unsupervised learning, the DNN model approximates a different unknown function: the one that models the relationship between the input of the model and power control coefficients to directly maximize the minimum SE. The unsupervised learning method, which is used in this paper, also reduces the need for large amounts of labeled data.

B. Essential Inputs for Power Control Optimization

Efficient function approximation using DNN necessitates a thorough understanding of the structural complexities in \mathbf{B} and the interdependencies between each user's SE and key variables, including pilot contamination dynamics and the matrices \mathbf{B} and \mathbf{M} .

Consider the following two terms contributing to $2K^2$ variables, with indices i and k ranging from 1 to K :

$$\bar{\boldsymbol{\mu}}_i^T \boldsymbol{\nu}_{ik} = \left| \psi_k^T \psi_i^* \right| \sum_{m=1}^M \sqrt{\nu_{mi}} \mu_{mi} \frac{\beta_{mk}}{\beta_{mi}} \quad (5)$$

and

$$\|\mathbf{D}_k \bar{\boldsymbol{\mu}}_i\|^2 = \sum_{m=1}^M \mu_{mi}^2 \beta_{mk}. \quad (6)$$

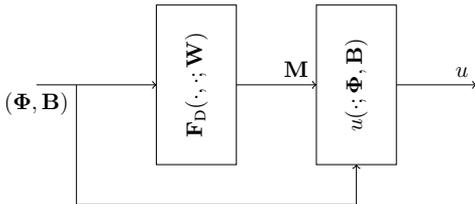


Fig. 2: Diagram of the unsupervised learning framework for downlink power control in CFmMIMO. The DNN, $F_D(\cdot, \cdot; \mathbf{W})$, takes the large-scale fading coefficient matrix \mathbf{B} and the pilot allocation information Φ as inputs to generate the power control coefficients. The DNN is trained to generate these coefficients to minimize the average of the utility function, $u(\cdot; \Phi, \mathbf{B})$.

From (1) and (2), the SE expressions for all users are composed of these $2K^2$ variables. Thus, their interplay determines $\min_{1 \leq k \leq K} \text{SE}_k$. To maximize $\min_{1 \leq k \leq K} \text{SE}_k$, the DNN should learn to generate the power control coefficient matrix \mathbf{M} , which determines the $2K^2$ control variables directly influencing the minimum SE.

Let $\Phi \in \mathbb{R}^{K \times K}$ represent the pilot allocation information, where $|\psi_i^T \psi_j^*|^2$ is the element at (i, j) of Φ , i.e., $\phi_{i,j}$. From (5), incorporating Φ into the DNN is essential. Without it, the DNN cannot generate \mathbf{M} that can effectively control the K^2 variables associated with $\bar{\mu}_i^T \nu_{ik}$. Similarly, equations (5) and (6) underscore the importance of including \mathbf{B} as input, while equation (5) emphasizes the need to preserve inter-user relationships between its columns, especially the element-wise ratios. Therefore, the DNN should incorporate both Φ and \mathbf{B} as inputs while preserving the structure of \mathbf{B} .

C. Optimization Problem for DNN Training

Assuming pilot allocation is performed prior to power control, and Φ is available as input to the power control algorithm, define an arbitrary DNN, $F_D(\cdot, \cdot; \mathbf{W})$, that takes Φ and \mathbf{B} as inputs, $\mathbf{M} = F_D(\Phi, \mathbf{B}; \mathbf{W})$, where \mathbf{W} represents the trainable parameters.

Since Φ is known, unlike in Sec. II, the SE is now a function solely of power control coefficients, while the large-scale fading coefficients and pilots are considered fixed parameters that characterize the function.

Assuming fixed BS placement and a path loss model, generate a set of large-scale fading coefficient matrices $\{\mathbf{B}^{[p]} \in \mathbb{R}_+^{M \times K} \mid p = 1, \dots, P\}$ for P random user placements, which are then used by the pilot allocation algorithm to produce $\{\Phi^{[p]} \in \mathbb{R}_+^{M \times K} \mid p = 1, \dots, P\}$.

Similar to the APG method in [5], consider a soft-minimum utility function with smoothening parameter λ as:

$$u(\mathbf{M}; \Phi, \mathbf{B}) = -\frac{1}{\lambda} \ln \left(\frac{1}{K} \sum_{k=1}^K e^{-\lambda \text{se}_k(\mathbf{M}; \Phi, \mathbf{B})} \right), \quad (7)$$

and the utility function for each sample as

$$u^{[p]}(\mathbf{W}) = u \left(F_D(\Phi^{[p]}, \mathbf{B}^{[p]}; \mathbf{W}); \Phi^{[p]}, \mathbf{B}^{[p]} \right). \quad (8)$$

The DNN training finds the optimal weights \mathbf{W}_{opt} :

$$\mathbf{W}_{\text{opt}} = \underset{\mathbf{W}}{\text{argmax}} \frac{1}{P} \sum_{p=1}^P u^{[p]}(\mathbf{W}). \quad (9)$$

During inference, this aims to replace expensive solvers for the max-min fairness problem, seeking the mapping $F_D(\Phi, \mathbf{B}; \mathbf{W}_{\text{opt}})$ that achieves similar performance. The unsupervised learning setup is depicted by Fig. 2.

D. Towards an Efficient DNN Design

The structure of the DNN, $F_D(\cdot, \cdot; \mathbf{W})$, plays a crucial role in achieving comparable performance to iterative solvers. Its design is essential for effectively handling the function approximation task. While FCN models are favored for their universal approximation abilities, they require flattened \mathbf{B} , leading to convergence issues in large-scale systems. They also fail to incorporate Φ , limiting their use to smaller systems. These limitations underscore the need for alternative architectures that preserve structural integrity of \mathbf{B} and effectively address pilot contamination, enabling scalability.

IV. THE PAPC TRANSFORMER

A. Overview of GPT

Introduced in [40], the transformer architecture has significantly impacted the field of natural language processing (NLP), enabling advancements in tasks such as machine translation, text summarization, and sentiment analysis. Among the architectures inspired by this model, generative pre-transformers (GPT) [41] stand out for their impact on language generation. While the proposed neural network differs in both purpose and training methodology from GPT, it shares significant structural similarities. This subsection provides a broad overview of GPT to lay the foundation for the proposed network's design.

GPT's preprocessing involves tokenization, embedding, and positional encoding. The input text is broken into tokens (words or subwords), transformed into numerical vectors through embedding to capture semantic features, and then augmented with positional encoding to preserve the sequence order. These encoded vectors are fed into GPT, which uses its attention mechanisms to extract context-aware features and perform tasks like next-token prediction.

At the core of GPT is the masked multi-head attention (MMHA) mechanism, which captures dynamic *interrelationships* among tokens by focusing on relevant parts of the input sequence. Masking enforces causality, restricting the model's attention to the current and preceding tokens, supporting GPT's autoregressive nature. The multi-head structure allows simultaneous exploration of token dependencies from various perspectives, enhancing the model's ability to generate coherent text.

The GPT architecture consists of multiple blocks, each integrating MMHA with normalization layers and feed-forward networks. These blocks sequentially process the input, refining token representations to predict the next token accurately.

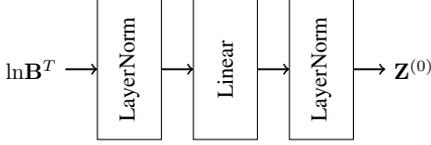


Fig. 3: Preprocessing stage of the CFmMIMO power control model. The input matrix \mathbf{B} is transposed, log-transformed, and mapped to a higher-dimensional representation $\mathbf{Z}^{(0)}$ for use in subsequent transformer blocks.

Although GPT was originally developed for language tasks, its key features—attention and masking—can be adapted to non-NLP applications such as power control. In this context, the attention mechanism can model inter-user relationships from input data like \mathbf{B} , and the masking mechanism can be adjusted to incorporate Φ as input.

B. PAPC for Downlink Power Control

This paper proposes the PAPC transformer model, designed to address the challenges of DNN-based downlink power control in CFmMIMO. Building upon the foundational principles of transformer architectures like GPT, PAPC effectively models the interactions between the columns of the large-scale fading matrix \mathbf{B} , representing the dynamic inter-user channel dependencies. By learning these interdependencies, the model computes power control coefficients. It incorporates pilot allocation matrix Φ using customized masking functionality. The PAPC model and its core components are outlined here.

1) *Layer Normalization*: A customized layer normalization technique is introduced as a fundamental unit of the PAPC transformer, applied at several stages throughout the architecture.

Layer normalization consists of two key steps: normalization and feature-specific scaling and shifting. Let \mathbf{C} represent the input matrix, where each row is a feature vector of length \bar{M} corresponding to a user.

First, the normalization process computes the mean and standard deviation across all elements of \mathbf{C} . Each element is then adjusted by subtracting the mean and dividing by the standard deviation, resulting in $\bar{\mathbf{C}}$.

Next, feature-specific scaling and shifting are applied to the normalized matrix $\bar{\mathbf{C}}$ using the trainable vectors α and β , both of length \bar{M} :

$$\text{LayerNorm}(\mathbf{C}; \mathbf{W}_L) = (\mathbf{1}_K \alpha^T) \odot \bar{\mathbf{C}} + \mathbf{1}_K \beta^T,$$

where $\mathbf{W}_L = \{\alpha, \beta\}$ represents the trainable parameters.

Unlike in GPT, where each input feature vector is normalized independently, the PAPC transformer normalizes all feature vectors together, preserving inter-user relationships. Additionally, it applies feature-specific scaling and shifting for greater flexibility across feature dimensions, unlike GPT's scalar-based approach.

2) *Preprocessing Stage*: The preprocessing stage of the PAPC transformer is designed to prepare the input large-scale fading coefficients matrix \mathbf{B} for use within the transformer architecture. The matrix is first transposed so that each row

corresponds to a user, allowing the model to treat each user's fading coefficients as a single unit. All the linear layers in the transformer architecture are thus performed row-wise.

To handle the significant variation in the values of \mathbf{B} , an element-wise logarithm is applied. Each row of $\ln(\mathbf{B}^T)$ is linearly mapped into a higher-dimensional space of length \bar{M} ($\bar{M} > M$), extracting a richer set of features. Layer normalization is applied before and after the mapping to ensure stable training. Let \mathbf{W}_P represent all the trainable parameters in this preprocessing step, then the output is denoted as $\mathbf{Z}^{(0)} = \mathbf{F}_P(\mathbf{B}; \mathbf{W}_P)$. Fig 3 depicts this functionality.

In NLP models like GPT, preprocessing involves embedding discrete tokens into continuous vectors and applying positional encoding. In the PAPC transformer, each user's fading coefficients are treated similarly to tokens, with a linear mapping analogous to the embedding step, transforming each user's vector into a higher-dimensional space. Semantic similarity between tokens can be compared to channel similarities between users. Since the user order is irrelevant, positional encoding is omitted. The resulting matrix, $\mathbf{Z}^{(0)}$, provides a rich representation of the propagation environment, similar to the context in NLP applications.

The primary computational cost of a single forward pass comes from the linear mapping, with complexity $\mathcal{O}(\bar{M}MK)$.

3) *MMHA*: The MMHA module is the core building block of the PAPC transformer. It processes input data across multiple attention heads and applies masking to capture dynamic inter-user relationships.

The MMHA module takes the matrix $\mathbf{X} \in \mathbb{R}^{K \times \bar{M}}$ and the mask matrix $\Phi \in \mathbb{R}^{K \times K}$ as inputs and produces $\mathbf{Y} \in \mathbb{R}^{K \times \bar{M}}$. The module operates with H attention heads, where \bar{M} is an integer multiple of H ($\bar{M} = HD$; H and D are integers). The overall architecture of the MMHA module is illustrated in Fig. 4.

As shown in Fig. 4a, for each attention head h , $h = 1, \dots, H$ the input matrix \mathbf{X} is transformed into three matrices: Query ($\mathbf{Q}^{(h)}$), Key ($\mathbf{K}^{(h)}$), and Value ($\mathbf{V}^{(h)}$), each of size $K \times D$, through separate linear transformations. Using these three matrices as input to a scaled dot-product self-attention mechanism, the head then computes the output matrix $\mathbf{Y}^{(h)}$.

The scaled dot-product self-attention mechanism, shown in Fig. 4b, begins with the computation of attention scores:

$$\mathbf{S}^{(h)} = \mathbf{Q}^{(h)} (\mathbf{K}^{(h)})^T / \sqrt{D},$$

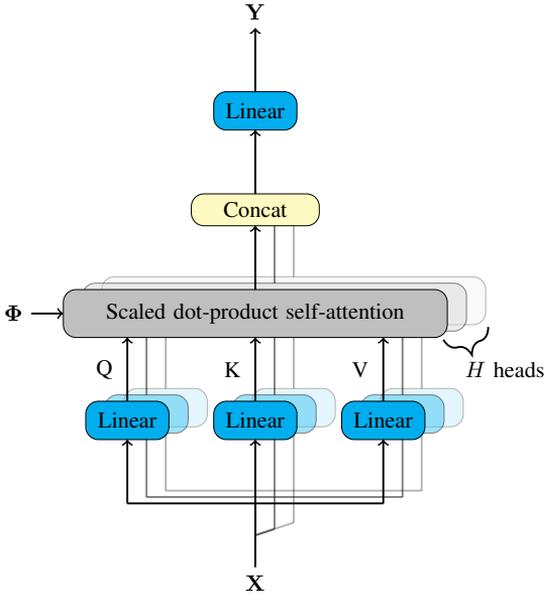
followed by a masking operation:

$$\bar{\mathbf{S}}^{(h)} = \mathbf{S}^{(h)} \odot \Phi.$$

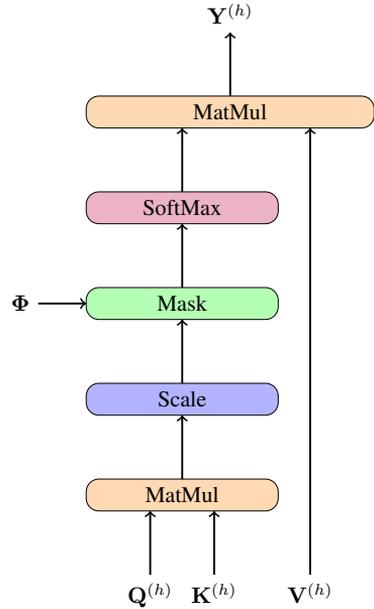
The softmax function [42] is then applied row-wise to $\bar{\mathbf{S}}^{(h)}$ to generate the attention weights $\mathbf{A}^{(h)}$, and compute

$$\mathbf{Y}^{(h)} = \mathbf{A}^{(h)} \mathbf{V}^{(h)}.$$

As shown in Fig. 4, the outputs from all H heads are concatenated along the feature dimension and passed through a linear layer to produce the final output \mathbf{Y} , compactly represented as $\mathbf{Y} = \mathbf{F}_M(\Phi, \mathbf{X}; \mathbf{W}_M)$, where \mathbf{W}_M denotes all the trainable parameters.



(a) Overview of MMHA with multi-head attention and output combination.



(b) Scaled dot-product self-attention with custom masking.

Fig. 4: MMHA architecture in the PAPC transformer, processing the input through multiple attention heads combined with masking feature to model inter-user relationships and handle pilot contamination.

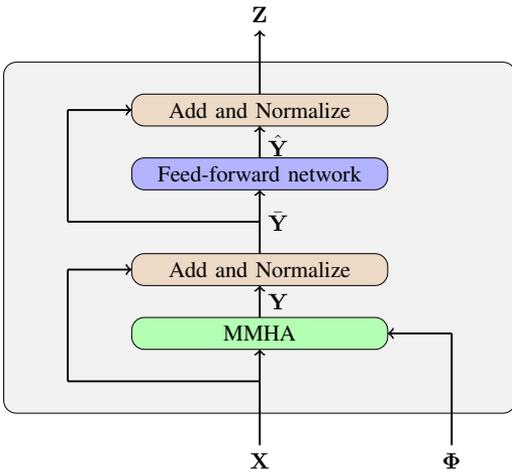


Fig. 5: The PAPC transformer block processes the input using an MMHA and a feed-forward network, with residual connections and layer normalizations.

In an attention head, each user's output is computed as a weighted combination of the feature vectors (rows) of the Value matrix, with the contribution of each user's input features determined by data-driven attention weights. Rather than manually crafting input features using element-wise ratios of large-scale fading coefficients, as described in Section III-B,

these heads learn to capture the inter-user relationships from the input data in the form of attention weights. These weights are computed independently across multiple heads, allowing the model to capture diverse relationships between users.

While GPT employs unidirectional attention for sequence generation—where each token is influenced only by its preceding tokens—this is achieved through binary masking, setting attention scores to $-\infty$ for future tokens to ensure their attention weights are zero after the softmax. In contrast, the PAPC transformer requires bidirectional attention, similar to bidirectional encoder representations from transformers (BERT) [43], to compute all rows of \mathbf{Y} simultaneously, thereby capturing inter-user relationships without sequence dependency. Here, masking is not employed for controlling directionality but rather for incorporating the matrix Φ , which reflects pilot allocation information, whose elements range from 0 to 1. The MMHA in PAPC manages user relationships and pilot contamination by integrating BERT's bidirectional attention with GPT-inspired masking.

In GPT, the MMHA module sets masked positions of attention scores to $-\infty$, whereas in the MMHA of PAPC, zeros in Φ lead to zeros in attention scores while resulting in nonzero attention weights. Intuitively, a zero attention weight is desirable for mutually uncontaminated pair of users, as channel similarities among them is irrelevant. This counter-intuitive design, while unexpected, has proven effective based on extensive simulations.

The dominant cost during a single forward pass of MMHA arises from computing the Query, Key, and Value matrices,

characterized by $\mathcal{O}(\bar{M}^2K)$.

4) *The P_{APC} Transformer Block*: The P_{APC} transformer block processes the input matrix \mathbf{X} and the mask matrix Φ , producing the output matrix \mathbf{Z} through attention and feed-forward operations, as shown in Fig. 5.

First, the MMHA module computes the intermediate matrix \mathbf{Y} as $\mathbf{F}_M(\Phi, \mathbf{X}; \mathbf{W}_M)$, where \mathbf{W}_M represents its trainable parameters. A residual connection then adds the input \mathbf{X} to \mathbf{Y} , and the result is normalized using layer normalization with trainable parameters \mathbf{W}_{T_1} , yielding

$$\bar{\mathbf{Y}} = \text{LayerNorm}(\mathbf{X} + \mathbf{Y}; \mathbf{W}_{T_1}).$$

Next, $\bar{\mathbf{Y}}$ is passed through a feed-forward (FF) network with one hidden layer, consisting of \bar{M} units with rectified linear unit (ReLU) activation [42]. It computes $\hat{\mathbf{Y}} = \text{FF}(\bar{\mathbf{Y}}; \mathbf{W}_{FF})$ with parameters \mathbf{W}_{FF} . Another residual connection adds $\bar{\mathbf{Y}}$ to $\hat{\mathbf{Y}}$, followed by layer normalization with parameters \mathbf{W}_{T_2} to produce

$$\mathbf{Z} = \text{LayerNorm}(\bar{\mathbf{Y}} + \hat{\mathbf{Y}}; \mathbf{W}_{T_2}).$$

Distinct from the GPT architecture, the P_{APC} transformer block incorporates the external mask matrix Φ as an input to MMHA.

The overall functionality of the P_{APC} transformer block is compactly represented as $\mathbf{Z} = \mathbf{F}_T(\Phi, \mathbf{X}; \mathbf{W}_T)$, where \mathbf{W}_T denotes all the trainable parameters. The computational complexity of the P_{APC} transformer block is dominated by the MMHA and FF modules, with a complexity of $\mathcal{O}(\bar{M}^2K)$ per forward pass.

5) *Postprocessing Stage*: The postprocessing stage is designed for the P_{APC} transformer, consisting of L transformer blocks that output $\mathbf{Z}^{(L)}$ with dimensions $K \times \bar{M}$. The goal is to convert $\mathbf{Z}^{(L)}$ into power control coefficients \mathbf{M} , using the additional input Φ , as shown in Fig. 6.

A linear mapping to M dimensions on each row of $\mathbf{Z}^{(L)}$ is applied:

$$\hat{\mathbf{M}} = \text{Linear}(\mathbf{Z}^{(L)}; \mathbf{W}_{O_1}),$$

where \mathbf{W}_{O_1} represents the trainable parameters. Let \mathbf{W}_{O_2} denote the trainable parameters of a subsequent normalization stage:

$$\bar{\mathbf{M}} = \text{LayerNorm}(\hat{\mathbf{M}}; \mathbf{W}_{O_2}).$$

Next, the matrix $\bar{\mathbf{M}}$ undergoes the following transformation:

$$\tilde{\mathbf{M}} = e^{-\text{ReLU}(\bar{\mathbf{M}}^T + 6)}$$

to ensure that each element of $\tilde{\mathbf{M}}$ remains bounded between 0 and 1, and initialized as small positive number.

The matrix $\tilde{\mathbf{M}}$ is then multiplied by the diagonalized version of Φ , represented as $\hat{\Phi} = \text{diagonalize}(\Phi)$, where the off-diagonal elements of Φ are set to zero. The final output is obtained by projecting the result onto \mathcal{S} :

$$\mathbf{M} = \text{Proj}_{\mathcal{S}}(\tilde{\mathbf{M}}\hat{\Phi}). \quad (10)$$

The projection operation $\text{Proj}_{\mathcal{S}}(\cdot)$ is performed as a series of per-BS projections, as outlined in [12].

Let \mathbf{W}_O represent all the trainable parameters in the postprocessing stage, then, the overall functionality can be represented as $\mathbf{M} = \mathbf{F}_O(\Phi, \mathbf{Z}^{(L)}; \mathbf{W}_O)$.

In summary, the linear mapping reduces the dimensionality to match the number of base stations M , while the layer normalization ensures stability. The matrix transpose is performed to align with the required $M \times K$ structure of the final output. The ReLU and exponentiation ensure that the elements of $\tilde{\mathbf{M}}$ are in $[0, 1]$, while the scalar shift of 6 initializes these elements to small positive numbers. The projection operation ensures that the final power control values satisfy the necessary constraints. Finally, the multiplication of $\tilde{\mathbf{M}}$ with the diagonalized Φ enables the P_{APC} transformer to flexibly handle a varying number of users avoiding a redesign.

To handle varying user counts (varying K feature), the system supports up to K_{MAX} users by padding. When $K < K_{\text{MAX}}$, \mathbf{B} is padded with a small constant (e.g., $6 \cdot 10^{-13}$) and Φ with zeros. Combined with the matrix multiplication of $\tilde{\mathbf{M}}$ and diagonalized Φ , this ensures that outputs for users beyond K remain zero, allowing the system, designed for K_{MAX} users, to adapt without a redesign.

The computational complexity of a forward pass is dominated by the linear mapping layer, which has a complexity of $\mathcal{O}(\bar{M}MK)$.

6) *P_{APC} Transformer Model Design*: The P_{APC} transformer model consists of three main components: the input preprocessing stage, a sequence of L P_{APC} transformer blocks, and the postprocessing stage. It takes \mathbf{B} as input and produces the power control coefficient matrix \mathbf{M} . Fig 7 represents the P_{APC} architecture.

As described in Section IV-B2, the preprocessing stage generates $\mathbf{Z}^{(0)}$ taking \mathbf{B} as input. For l ranging from 1 to L , the output of each transformer block is given by $\mathbf{Z}^{(l)} = \mathbf{F}_T(\Phi, \mathbf{Z}^{(l-1)}; \mathbf{W}_T^l)$, where \mathbf{W}_T^l is the trainable parameters of the l^{th} block. The postprocessing stage takes Φ and $\mathbf{Z}^{(L)}$ as input to produce \mathbf{M} . Note that all transformer blocks, as well as the postprocessing stage, use the same mask matrix Φ as an additional input.

Assuming $L \ll M$ and \bar{M} is of order M , as observed from the simulations, the computational complexity for a single forward pass in the P_{APC} transformer is $\mathcal{O}(M^2K)$. In contrast, the computational complexity of the APG is $\mathcal{O}(M_1MK^2)$ [12], where M_1 is the number of iterations and is comparable to M . This comparison underscores the substantial computational advantage of the learning-based P_{APC} transformer over the APG.

The P_{APC} transformer is compactly represented as

$$\mathbf{M} = \mathbf{F}_{\text{PAPC}}(\Phi, \mathbf{B}; \mathbf{W}_{\text{PAPC}}),$$

with \mathbf{W}_{PAPC} as the trainable parameters. The transformer \mathbf{F}_{PAPC} is trained to approximate the optimal mapping from the large-scale fading matrix and the pilot allocation information matrix to power control coefficients that maximize the empirical average of smoothed minimum SE as given in (9). By leveraging attention mechanism to capture structural relationships in \mathbf{B} and incorporating Φ , it offers a scalable and efficient solution for downlink power control in CFmMIMO networks.

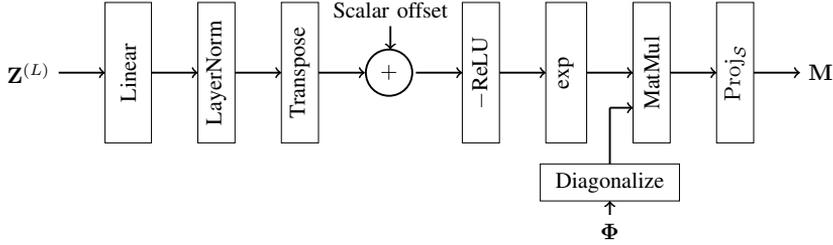


Fig. 6: Postprocessing stage of the PAPC, converting the final transformer block’s output into power control coefficients through a series of techniques to ensure necessary constraints and to support varying K feature.

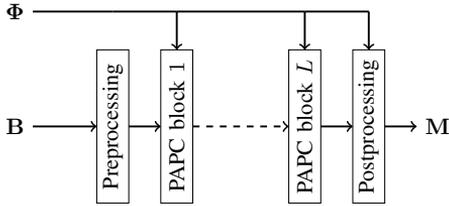


Fig. 7: PAPC transformer architecture using L PAPC transformer blocks. All the blocks and the postprocessing stage incorporate Φ .

7) *Training the DNN*: The PAPC is trained using the PyTorch library, which automatically handles the backpropagation.³ The adaptive moment estimation (ADAM) optimizer is employed, configured with parameters $\beta_1^{\text{ADAM}} = 0.9$, $\beta_2^{\text{ADAM}} = 0.98$, and $\epsilon^{\text{ADAM}} = 10^{-9}$, as suggested in [40]. Training is conducted over 16 epochs with mini-batches of size 1024. A learning rate scheduler adjusts the learning rate (l_{rate}) based on the training step number n_{step} , as in [40]:

$$l_{\text{rate}} = d_{\text{mod}}^{-0.5} \cdot \min(n_{\text{step}}^{-0.5}, n_{\text{step}} \cdot n_{\text{warmup}}^{-1.5}) \quad (11)$$

where n_{warmup} is set to 4000, and d_{mod} depends on the model’s size.⁴

V. NUMERICAL RESULTS

A. Simulation Setup

The performance of the PAPC transformer in downlink power control is evaluated in a CFmMIMO system with a density of 1000 BSs per sq. km. Various scenarios with different numbers of BSs and users, where each BS is equipped with $N = 4$ antennas, is considered. A wrap-around topology is assumed to simulate a large area to avoid boundary effects.

The distance between the m^{th} BS and the k^{th} user is d_{mk} [km]. A three-slope path loss model defines the path loss:

³We thank Andrei Palshin, Department of Information and Communications Engineering, Aalto University, for his invaluable support in training, fine-tuning, and validating the deep learning models used in this work.

⁴In Section V, d_{mod} is set according to the simulation scenarios considered. The complete implementation is available in the GitHub repository [44].

| | |
|--|----------------------|
| BS Density | 1000 APs per sq. km. |
| Length of the coherence block | 200 symbols |
| Length of the pilot sequence | 20 symbols |
| L_0 | 140.72 dB |
| d_0 | 0.01 km |
| d_1 | 0.05 km |
| Standard deviation of shadow fading (σ_{sh}) | 8 dB |
| Noise power spectral density N_0 | -173.98 dBm/Hz |
| BandWidth | 20 MHz |
| Total Noise power at the receiver (P_n) | -91.97 dBm |
| Transmit SNR of uplink pilot (ζ_p) | $1/P_n$ |
| Transmit SNR of downlink data (ζ_d) | $0.2/P_n$ |
| Smoothing parameter (λ) | 3 |

TABLE I: Simulation Setup Parameters

$PL_{mk} = -L_0 - 15 \log_{10}(d_1) - 20 \log_{10}(d'_{mk})$ [dB], where d'_{mk} is

$$d'_{mk} = \begin{cases} d_0 & d_{mk} \leq d_0 \\ d_{mk} & d_0 < d_{mk} \leq d_1 \\ d_1 & d_{mk} > d_1. \end{cases}$$

The large-scale fading coefficient of the corresponding channel is $\beta_{mk} = PL_{mk} + z_{mk}$ [dB], with $z_{mk} \sim \mathcal{N}(0, \sigma_{\text{sh}}^2)$ accounting for shadow fading. The parameters are set as $L_0 = 140.72$ dB, $d_0 = 0.01$ km, $d_1 = 0.05$ km, and $\sigma_{\text{sh}} = 8$ dB, following [5].

The noise power is $P_n = BW10^{(N_0 + N_f - 30)/10}$ [W]. Assuming a noise figure of $N_f = 9$ dB, the noise power spectral density $N_0 = -173.98$ dBm/Hz, and a channel bandwidth of $BW = 20$ MHz, the transmit SNR for the uplink pilot and downlink data are $\zeta_p = 0.2/P_n$ and $\zeta_d = 1/P_n$, respectively.

The coherence block and the pilot sequence lengths are $\tau = 200$ symbols and $\tau_p = 20$ symbols, respectively. The pilot allocation method assigns pilots from τ_p orthogonal sequences, giving the first $\min(K, \tau_p)$ users unique pilots, and then randomly selecting/reusing pilots for the remaining users if $K > \tau_p$. The smoothing parameter in (7) is set as $\lambda = 3$. Table I summarizes the simulation setup.

B. Neural Network Models and Training

To demonstrate the potential of PAPC, an FCN model is trained alongside PAPC for various scenarios. In all the scenarios, 2000 samples are used for evaluation, while they are trained on $P = 12,000,000$ samples unless stated otherwise.

| Scenario | M | K_{MIN} | K_{MAX} | \hat{M} | \hat{M} | d_{mod} |
|----------|-----|------------------|------------------|-----------|-----------|------------------|
| 0 | 10 | — | 4 | 80 | 160 | 16 |
| 1 | 100 | — | 20 | 500 | 1000 | 100 |
| 2 | 100 | — | 40 | 500 | 571 | 100 |
| 3 | 100 | 40 | 80 | 500 | — | 100 |

TABLE II: Hyperparameters for Each Scenario

Training occurs on Graphics processing units (GPUs)⁵, and testing is done without GPU assistance to ensure fair comparison of computational times across algorithms.

The FCN model uses a flattened \mathbf{B} , that becomes a vector of length MK , as its input. This is transformed through a layer normalization stage before passing it through three fully connected linear layers, including an input, a hidden, and an output layer. The input and hidden layers are followed by a corresponding layer normalization module and a ReLU unit. The number of features in the hidden layer is \hat{M} . Furthermore, a matrix reshaping is performed, followed by a postprocessing operation similar to PAPC's postprocessing module, but without the matrix multiplication used in PAPC to handle the varying K feature.

C. Scenarios and Evaluation Strategy

The performance of PAPC is compared against FCN, a simple equal power allocation (EPA) algorithm, and the APG algorithm. In EPA, each BS assigns equal power to all the users in the downlink signal. The empirical cumulative distribution function (CDF) of the per-user SE is used to represent the performance curve of all algorithms.

Four distinct scenarios are examined, ranging from Scenario 0 to Scenario 3. Scenario 0 represents a small-scale CFmMIMO network with $M = 10$ BSs and $K = 4$ users within an area of 0.01 sq. kms. In contrast, Scenarios 1, 2, and 3 expand the network to $M = 100$ BSs, with user counts of $K = 20$, $K = 40$, and $K = 80$, respectively, across an area of 0.1 sq. kms. For all the scenarios, the number of transformer blocks is $L = 3$ and the number of heads is $H = 5$.

The use of PAPC enables the extension of CFmMIMO from the small-scale Scenario 0, consistent with similar ranges of network sizes discussed in existing literature, to larger configurations considered in Scenarios 1 through 3. In Scenarios 2 and 3, it is important to note that pilot reuse leads to pilot contamination.

To compare the performance of FCN and PAPC, \hat{M} is set in such a way that the number of trainable parameters in both the networks is approximately the same. Due to heavy computational requirements and poor performance, the evaluation of FCN is omitted in Scenario 3. To demonstrate the varying K feature of PAPC, the model for Scenario 3 is trained for the values of K between K_{MIN} and K_{MAX} . Table II summarizes hyperparameters for each scenario.

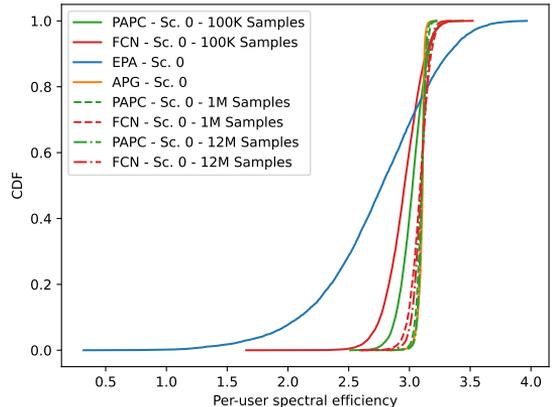


Fig. 8: CDF comparison of PAPC, FCN, EPA, and APG in Scenario 0 for different training samples. To interpret the results, note that for the max-min fairness objective, a CDF curve that ascends sharply and is also positioned further to the right compared to other curves is considered advantageous. Thus, PAPC outperforms FCN and EPA, approaching APG's performance faster as the number of samples increases.

D. Validation of PAPC in Contamination-Free Small-Scale CFmMIMO

Scenario 0 is a small and contamination-free CFmMIMO system used to assess the PAPC's performance. The FCN and PAPC models are trained with $P = 100,000$, $P = 1,000,000$, and $P = 12,000,000$ samples. For this scenario, Fig. 8 compares the CDFs of PAPC with FCN, EPA, and APG.

In this simple and contamination-free scenario, both PAPC and FCN surpass the EPA method with as few as 100,000 training samples, highlighting the effectiveness of learning-based approaches even with limited data. As the number of training samples increases (from 100,000 to 1,000,000 and 12,000,000), both models continue to improve, with their CDF curves moving closer together, indicating a reduced performance gap with the benchmark, APG. PAPC, however, consistently reaches performance closer to APG faster than FCN, likely due to its structural design and effective utilization of pilot information (the fact that there is no contamination). With 12,000,000 samples, PAPC's performance closely aligns with that of the APG algorithm, demonstrating its ability to achieve results comparable to the benchmark with sufficient training.

E. PAPC Performance in Large-Scale CFmMIMO Scenarios

Scenarios 1 to 3 represent large-scale CFmMIMO systems. Scenario 1 is contamination-free, while Scenarios 2 and 3 involve pilot contamination, with Scenario 3 experiencing heavy contamination.

The FCN and PAPC models are trained for Scenario 1 and Scenario 2. Additionally, the PAPC is trained for Scenario 3

⁵We acknowledge the computational resources provided by the Aalto Science-IT project.

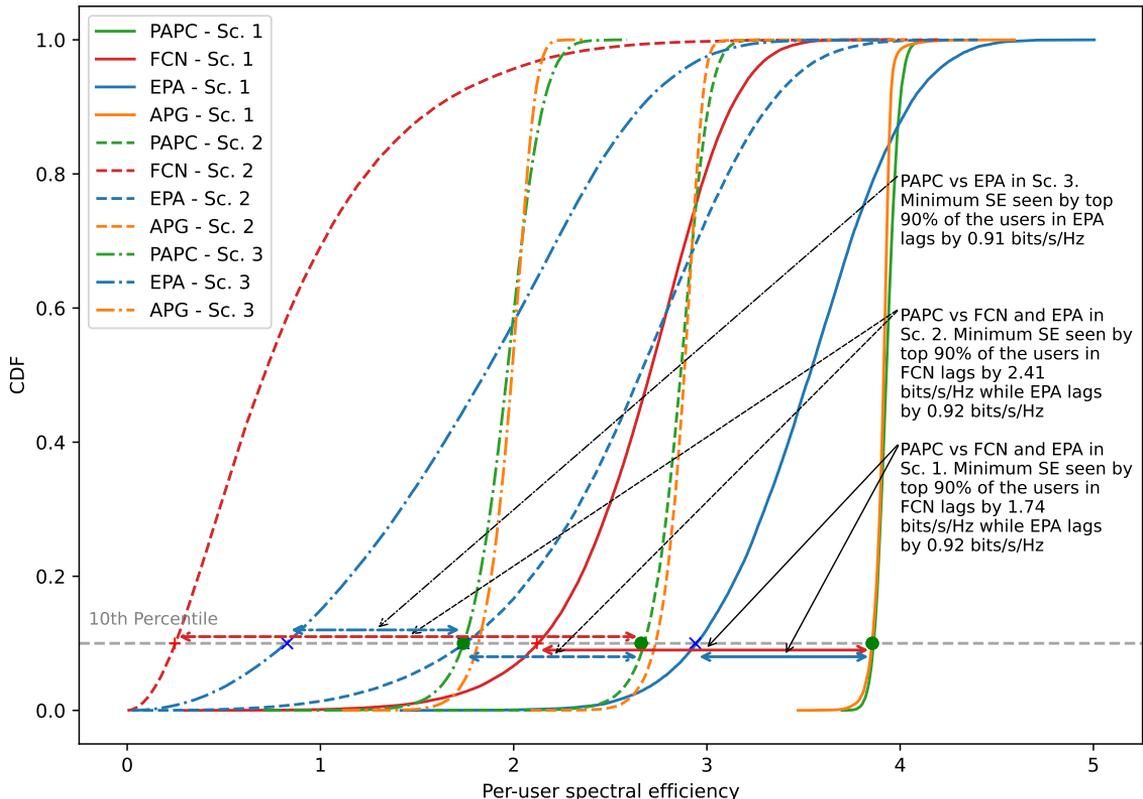


Fig. 9: CDF comparison across Scenarios 1 to 3 for different algorithms. PAPC consistently approaches APG performance, outperforming other algorithms due to its masking and attention mechanisms. FCN struggles due to its lack of structure and pilot allocation information.

by enabling the varying K feature. While testing this scenario, the input samples are generated using a fixed number of users, $K = K_{\text{MAX}}$. Fig. 9 provides a comparative analysis between the CDFs of PAPC, FCN, EPA, and APG.

PAPC consistently approaches the performance of APG in all scenarios, while FCN struggles due to structural inefficiencies, such as the input flattening operation. In Scenario 2, pilot contamination further worsens FCN's performance as it lacks pilot allocation information, making it perform worse than both PAPC and EPA. In contrast, PAPC effectively handles pilot contamination, leveraging its masking feature to maintain robust performance.

A quantitative analysis of the minimum SE observed by the top 90% of users further demonstrates PAPC's superiority. In Scenario 1, FCN lags behind PAPC by 1.74 bits/s/Hz, and EPA lags by 0.92 bits/s/Hz. In Scenario 2, FCN falls behind PAPC by 2.41 bits/s/Hz, while EPA lags by 0.92 bits/s/Hz. Finally, in Scenario 3, where heavy contamination is present, EPA lags behind PAPC by 0.91 bits/s/Hz. Additionally, the minimum SE observed by PAPC lags behind APG by only 0.08 bits/s/Hz in both Scenario 2 and Scenario 3, while

for Scenario 1, the difference is negligible⁶. These results emphasize the robustness of PAPC across different levels of contamination, providing consistent gains compared to other methods.

Scenario 3 is the most critical one for evaluating PAPC, as it subjects the model to severe pilot contamination. While being trained with the varying K feature, PAPC shows strong test performance with a fixed number of users, effectively managing the contamination. Since Scenario 3 addresses the case of large-scale configuration, this scenario is used to compare the computational efficiency of PAPC, EPA, and APG. Table III details the average computational time for each approach, measured on a 64-bit Windows-10 system with 16 GB RAM and an Intel(R) Xeon(R) Platinum 8176, 2.10 GHz, without GPU testing.

From the table, the computationally inefficient EPA is the fastest algorithm, while PAPC achieves comparable performance to that of APG, but it is nearly 1000 times faster than APG.

⁶This detail is not annotated in the figure to avoid further visual congestion.

| Algorithm | Run-time (in secs) |
|-----------|--------------------|
| APG | 38.7373 |
| PAPC | 0.0262 |
| EPA | 0.0003 |

TABLE III: Run-time of the algorithms in Scenario 3. EPA is the fastest algorithm, while PAPC achieves comparable performance to that of APG, but it is nearly 1000 times faster than APG.

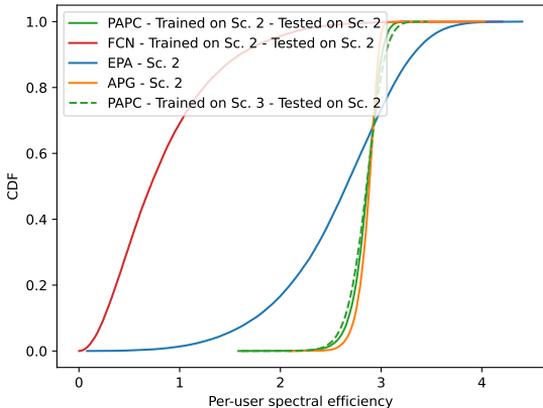


Fig. 10: Comparison of PAPCs trained on Scenario 2 and Scenario 3 (tested with $K = 40$), showing matching performance on Scenario 2 and validating that larger configurations with padding do not compromise results.

F. Evaluating the Flexibility of PAPC

Fig. 10 compares PAPC trained in Scenario 3 with varying K feature enabled, evaluated with fixed $K = 40$, to PAPC trained and tested in Scenario 2 with fixed $K = 40$. The performance of PAPC in both cases is nearly identical, validating that the padding and postprocessing tricks for handling varying K feature do not compromise the performance.

Fig. 11 presents the performance of PAPC when tested in Scenario 3 with the varying K feature enabled. The results show that PAPC maintains its strong performance, matching the APG algorithm and outperforming EPA, demonstrating its ability to dynamically adjust to fluctuating user counts without any loss in efficiency.

VI. CONCLUSION

The proposed PAPC transformer offers an innovative and efficient solution for downlink power control in CFmMIMO by utilizing the attention mechanism to leverage inter-user relationships and incorporating pilot allocation information via a novel masking technique. This enables PAPC to handle pilot contamination effectively, a limitation that confined prior learning-based methods to small-scale systems. By demonstrating scalability to a system size as large as $MK = 8000$ —the first in literature of learning-based solutions—PAPC significantly outperforms FCNs and matches the performance of traditional algorithms like APG with far greater computational efficiency. The PAPC’s ability to adapt

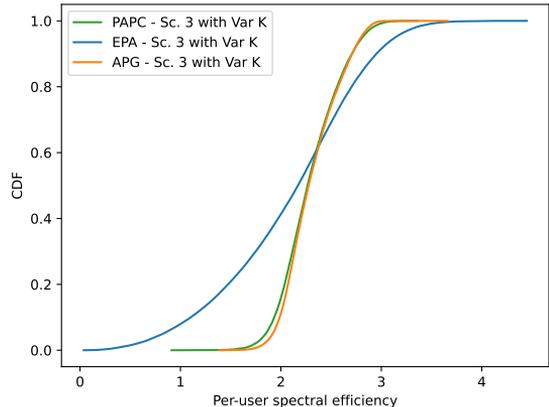


Fig. 11: PAPC performance in Scenario 3 with varying K , comparable to APG and surpassing EPA. PAPC maintains its strong performance, matching the APG algorithm and outperforming EPA, demonstrating its ability to dynamically adjust to fluctuating user counts without any loss in efficiency.

to varying numbers of users provides additional flexibility. With computational efficiency up to 1000 times faster than APG, PAPC offers a scalable solution for power control, with the potential to extend beyond the systems explored here, though scalability remains dependent on available training resources.

REFERENCES

- [1] R. W. Heath Jr, T. Wu, Y. H. Kwon, and A. C. K. Soong, “Multiuser MIMO in Distributed Antenna Systems With Out-of-Cell Interference,” *IEEE Trans. on Signal Process.*, vol. 59, no. 10, pp. 4885–4899, 2011.
- [2] R. Heath, S. Peters, Y. Wang, and J. Zhang, “A current perspective on distributed antenna systems for the downlink of cellular systems,” *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 161–167, 2013.
- [3] Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sankhikhi, “Wireless network cloud: Architecture and system requirements,” *IBM J. of Research and Development*, vol. 54, no. 1, pp. 4:1–4:12, 2010.
- [4] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, “Cloud radio access network (C-RAN): a primer,” *IEEE Network*, vol. 29, no. 1, pp. 35–41, 2015.
- [5] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, “Cell-Free Massive MIMO Versus Small Cells,” *IEEE Trans. on Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.
- [6] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, M. Debbah, and P. Xiao, “Max–Min Rate of Cell-Free Massive MIMO Uplink With Optimal Uniform Quantization,” *IEEE Trans. on Commun.*, vol. 67, no. 10, pp. 6796–6815, 2019.
- [7] E. Björnson and L. Sanguinetti, “Making Cell-Free Massive MIMO Competitive With MMSE Processing and Centralized Implementation,” *IEEE Trans. on Wireless Commun.*, vol. 19, no. 1, pp. 77–90, 2020.
- [8] O. T. Demir, E. Björnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO*, 2021.
- [9] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, “Ubiquitous cell-free massive MIMO communica-

- tions," *EURASIP J. on Wireless Commun. and Networking*, vol. 2019, no. 1, pp. 1–13, Dec 2019.
- [10] A. Burr, M. Bashar, and D. Maryopi, "Cooperative Access Networks: Optimum Fronthaul Quantization in Distributed Massive MIMO and Cloud RAN - Invited Paper," in *Proc. IEEE 87th Veh. Tech. Conf. (VTC Spring)*, 2018, pp. 1–5.
- [11] H. Masoumi and M. J. Emadi, "Performance Analysis of Cell-Free Massive MIMO System With Limited Fronthaul Capacity and Hardware Impairments," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 2, pp. 1038–1053, 2020.
- [12] M. Farooq, H. Q. Ngo, E. Hong, and L. Tran, "Utility Maximization for Large-Scale Cell-Free Massive MIMO Downlink," *IEEE Trans. on Commun.*, vol. 69, no. 10, pp. 7050–7062, 2021.
- [13] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy Efficiency in Cell-Free Massive MIMO with Zero-Forcing Precoding Design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, 2017.
- [14] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the Total Energy Efficiency of Cell-Free Massive MIMO," *IEEE Trans. on Green Commun. and Networking*, vol. 2, no. 1, pp. 25–39, 2018.
- [15] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink Training in Cell-Free Massive MIMO: A Blessing in Disguise," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 11, pp. 5153–5169, 2019.
- [16] E. Björnson and L. Sanguinetti, "Scalable Cell-Free Massive MIMO Systems," *IEEE Trans. on Commun.*, vol. 68, no. 7, pp. 4247–4261, 2020.
- [17] S. Buzzi and C. D'Andrea, "Cell-Free Massive MIMO: User-Centric Approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, 2017.
- [18] T. C. Mai, H. Q. Ngo, and L. Tran, "Energy Efficiency Maximization in Large-Scale Cell-Free Massive MIMO: A Projected Gradient Approach," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 8, pp. 6357–6371, 2022.
- [19] N. Rajapaksha, K. B. Shashika Manosha, N. Rajatheva, and M. Latva-Aho, "Deep Learning-based Power Control for Cell-Free Massive MIMO Networks," in *Proc. IEEE Int. Conf. on Commun.*, 2021, pp. 1–7.
- [20] N. Rajapaksha, K. B. S. Manosha, N. Rajatheva, and M. Latva-aho, "Unsupervised Learning-Based Joint Power Control and Fronthaul Capacity Allocation in Cell-Free Massive MIMO With Hardware Impairments," *IEEE Wireless Commun. Lett.*, vol. 12, no. 7, pp. 1159–1163, 2023.
- [21] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised-Learning Power Control for Cell-Free Wireless Systems," in *Proc. IEEE 30th Annual Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, 2019, pp. 1–5.
- [22] A. Mazhari Saray and A. Ebrahimi, "MAX-MIN Power Control of Cell Free Massive MIMO System employing Deep Learning," in *Proc. 4th West Asian Symp. on Optical and Millimeter-wave Wireless Commun. (WASOWC)*, 2022, pp. 1–4.
- [23] C. D'Andrea, A. Zappone, S. Buzzi, and M. Debbah, "Uplink Power Control in Cell-Free Massive MIMO via Deep Learning," in *Proc. IEEE 8th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process. (CAMSAP)*, 2019, pp. 554–558.
- [24] Y. Zhang, J. Zhang, Y. Jin, S. Buzzi, and B. Ai, "Deep Learning-Based Power Control for Uplink Cell-Free Massive MIMO Systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.
- [25] H. Bao, K. Xiong, R. Zhang, P. Fan, D. Niyato, and K. B. Letaief, "User Cooperation with Power Control for Federated Learning in CFmMIMO Networks," in *Proc. IEEE Int. Conf. on Commun. Workshops (ICC Workshops)*, 2023, pp. 122–127.
- [26] I. M. Braga, R. P. Antonoli, G. Fodor, Y. C. B. Silva, and W. C. Freitas, "Decentralized Joint Pilot and Data Power Control Based on Deep Reinforcement Learning for the Uplink of Cell-Free Systems," *IEEE Trans. on Veh. Tech.*, vol. 72, no. 1, pp. 957–972, 2023.
- [27] M. Rahmani, M. Bashar, M. J. Dehghani, P. Xiao, R. Tafazolli, and M. Debbah, "Deep Reinforcement Learning-based Power Allocation in Uplink Cell-Free Massive MIMO," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, 2022, pp. 459–464.
- [28] X. Zhang, M. Kaneko, V. An Le, and Y. Ji, "Deep Reinforcement Learning-based Uplink Power Control in Cell-Free Massive MIMO," in *Proc. IEEE 20th Consumer Commun. & Networking Conf. (CCNC)*, 2023, pp. 567–572.
- [29] D. Yu, H. Lee, S. Hong, and S. Park, "Learning Decentralized Power Control in Cell-Free Massive MIMO Networks," *IEEE Trans. on Veh. Tech.*, vol. 72, no. 7, pp. 9653–9658, 2023.
- [30] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised-Learning Power Allocation for the Cell-Free Downlink," in *Proc. IEEE Int. Conf. on Commun. Workshops (ICC Workshops)*, 2020, pp. 1–5.
- [31] L. Luo, J. Zhang, S. Chen, X. Zhang, B. Ai, and D. W. K. Ng, "Downlink Power Control for Cell-Free Massive MIMO With Deep Reinforcement Learning," *IEEE Trans. on Veh. Tech.*, vol. 71, no. 6, pp. 6772–6777, 2022.
- [32] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised Learning for C-RAN Power Control and Power Allocation," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 687–691, 2021.
- [33] L. Salaün and H. Yang, "Deep Learning Based Power Control for Cell-Free Massive MIMO with MRT," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 01–07.
- [34] C. Hao, T. T. Vu, H. Q. Ngo, M. N. Dao, X. Dang, C. Wang, and M. Matthaiou, "Joint User Association and Power Control for Cell-Free Massive MIMO," *IEEE Internet of Things J.*, pp. 1–1, 2024.
- [35] L. Salaün, H. Yang, S. Mishra, and C. S. Chen, "A GNN Approach for Cell-Free Massive MIMO," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2022, pp. 3053–3058.
- [36] M. Zaher, O. T. Demir, E. Björnson, and M. Petrova, "Learning-Based Downlink Power Allocation in Cell-Free Massive MIMO Systems," *IEEE Trans. on Wireless Commun.*, vol. 22, no. 1, pp. 174–188, 2023.
- [37] Y. Zhang, J. Zhang, S. Buzzi, H. Xiao, and B. Ai, "Unsupervised Deep Learning for Power Control of Cell-Free Massive MIMO Systems," *IEEE Trans. on Veh. Tech.*, vol. 72, no. 7, pp. 9585–9590, 2023.
- [38] S. Chakraborty and B. R. Manoj, "Power Allocation in a Cell-Free MIMO System using Reinforcement Learning-Based Approach," in *Proc. National Conf. on Commun. (NCC)*, 2023, pp. 1–6.
- [39] A. K. Kocharlakota, S. A. Vorobyov, and R. W. Heath, "Attention Neural Network for Downlink Cell-Free Massive MIMO Power Control," in *Proc. 56th Asilomar Conf. on Signals, Systems, and Computers*, 2022, pp. 738–743.
- [40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is All you Need," in *Proc. Advances in Neural Inf. Process. Systems*, vol. 30. Curran Associates, Inc., 2017.
- [41] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018, accessed: 2023-01-06.
- [42] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [43] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019.
- [44] A. K. Kocharlakota, S. A. Vorobyov, and R. W. Heath, "CFmMIMO PC," GitHub repository, 2024. [Online]. Available: https://github.com/kochark1/CFmMIMO_PC



ISBN 978-952-64-2198-8 (printed)
ISBN 978-952-64-2199-5 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Information and Communications Engineering
www.aalto.fi

**BUSINESS +
ECONOMY**

**ART +
DESIGN +
ARCHITECTURE**

**SCIENCE +
TECHNOLOGY**

CROSSOVER

**DOCTORAL
THESES**