FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
POSTGRADUATE PROGRAM IN ELECTRICAL ENGINEERING

Doctoral Thesis

Scalable AP Selection Strategies for User-Centric Cell-Free Massive MIMO Networks

TD: 10/2024

MARX MIGUEL MIRANDA DE FREITAS

UFPA / ITEC / PPGEE
Guama University Campus
Belém – Pará – Brasil
2024

FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
POSTGRADUATE PROGRAM IN ELECTRICAL ENGINEERING

MARX MIGUEL MIRANDA DE FREITAS

Scalable AP Selection Strategies for User-Centric Cell-Free Massive MIMO Networks

TD: 10/2024

UFPA / ITEC / PPGEE
Guama University Campus
Belém – Pará – Brasil
2024

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

F862 Freitas, Marx Miguel Miranda de.
Scalable AP Selection Strategies for User-Centric Cell-Free Massive MIMO Networks / Marx Miguel Miranda de Freitas. — 2024.
xxii, 104 f. : il. color.

Orientador(a): Prof. Dr. João Crisóstomo Weyl Albuquerque Costa
Coorientador(a): Prof. Dr. Daniel Benevides da Costa E Prof. Dr. Luca Valcarenghi
Tese (Doutorado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2024.

1. AP Selection. 2. Computational Complexity. 3. Inter-CPU Coordination. 4. Scalability. 5. User-Centric Cell-free Massive MIMO. I. Título.

CDD 384

MARX MIGUEL MIRANDA DE FREITAS

Scalable AP Selection Strategies for User-Centric Cell-Free Massive MIMO Networks

Thesis submitted to the Examination Board of the Postgraduate Program in Electrical Engineering at UFPA for attaining the Doctor of Philosophy Degree in Electrical Engineering, Area of Concentration in Telecommunications, Research Line in Applied Electromagnetism.

Advisor: Prof. Dr. João Crisóstomo Weyl Albuquerque Costa
Co-advisor: Prof. Dr. Daniel Benevides da Costa
Co-advisor: Prof. Dr. Luca Valcarenghi

UFPA / ITEC / PPGEE
Guama University Campus
Belém – Pará – Brasil
2024

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

“ESTRATÉGIAS ESCALÁVEIS DE SELEÇÃO DE PONTOS DE ACESSO EM REDES CELL-FREE MASSIVE MIMO CENTRADAS NOS USUÁRIOS”

AUTOR: **MARX MIGUEL MIRANDA DE FREITAS**

TESE DE DOUTORADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA ELÉTRICA NA ÁREA DE TELECOMUNICAÇÕES.

APROVADA EM: 06/06/2024

BANCA EXAMINADORA:

DocuSigned by:
João Crisóstomo Weyl Albuquerque Costa
EFE1BE0196F444C...

Prof. Dr. João Crisóstomo Weyl Albuquerque Costa
(Orientador – PPGEE/UFPA)

DocuSigned by:
Daniel Benevides da Costa
6F84E648A9C6412...

Prof. Dr. Daniel Benevides da Costa
(Coorientador - KFUPM)

DocuSigned by:
Luca Valcarenghi
4255D8AFE04445A...

Prof. Dr. Luca Valcarenghi
(Coorientador Externo - TeCIP/SSSA)

DocuSigned by:
Aldebaro Klautau Jr
52E923F2EB0F447...

Prof. Dr. Aldebaro Barreto da Rocha Klautau Júnior
(Avaliador Interno - PPGEE/UFPA)

DocuSigned by:
Gilvan Soares Borges
515081E365D24E9...

Prof. Dr. Gilvan Soares Borges
(Avaliador Interno - PPGEE/UFPA)

DocuSigned by:
André Mendes Cavalcante
E7CD59D948CE4D9...

Dr. André Mendes Cavalcante
(Avaliador Externo - ERICSSON)

DocuSigned by:
Ugo Silva Dias
6AADC5CCB28B4F3...

Prof. Dr. Ugo Silva Dias
(Avaliador Externo - FT/UNB)

VISTO:

DocuSigned by:
Diego Lisboa Cardoso
1C8C1AE22FB24A4...

Prof. Dr. Diego Lisboa Cardoso
(Coordenador do PPGEE/ITEC/UFPA)

This work is dedicated to Jesus Christ, to my family, and to all those who strive to disseminate scientific knowledge around the world.

Acknowledgements

Throughout this journey, I have had the opportunity to count on the support and inspiration of various people who encouraged and assisted me at crucial moments.

First and foremost, I would like to express my gratitude to God for the gift of life and for inspiring me during times of adversity. I am also profoundly grateful to my family for their encouragement and support throughout every stage of my academic journey. They provided me with all love and affection I needed and always believed in me, even when I did not believe in myself. Especially my father, Asarias Freitas, who is my inspiration both in academia and life, and my mother, Lucilene Miranda, for all the emotional support. Furthermore, I thank my sister, Valéria Freitas, for her companionship and friendship. I know I can always count on my family.

I am grateful to my advisor, Prof. Dr. João Costa, for guiding me since my undergraduate studies and providing me with various opportunities within UFPA. I would also like to express my gratitude to the entire Applied Electromagnetism Laboratory for being like a family to me. Especially to Roberto Rodrigues, Gilvan Soares, and our administrator Liane Barbosa, who are tireless in their work. In addition to them, I thank my friends Daynara Souza and André Lucas, my work partners since 2013.

I am also profoundly grateful to my co-supervisor, Prof. Dr. Daniel Benevides da Costa, and to Dr. André Cavalcante. They always saw potential in my research and consistently encouraged me to move forward. Their insightful feedback and guidance significantly shaped the direction of my research. I also express my gratitude to Prof. Dr. Aldebaro Klautau, whose tireless dedication to teaching and advancing technology in the Amazon region has been invaluable.

I also extend my gratitude to the entire Sant'Anna University and TeCIP Institute for welcoming me so well during my time in Italy. I would like to give special thanks to Prof. Dr. Luca Valcarenghi, who supervised me in Italy and made significant contributions to my research. He also assisted me during one of the most difficult moments of my life abroad. I would also like to thank the friends I made in Italy and at TeCIP, who have a special place in my heart. Unfortunately, space is limited and there is no way to mention everyone, but you were fundamental throughout the doctorate.

Finally, I would like to thank Ericsson and CNPq for funding the research.

"*But the path of the just is as the shining light,*
that shineth more and more unto the perfect day."
Holy Bible. Proverbs 4:18 (KJV).

Resumo

Sistemas *user-centric* (UC) *cell-free* (CF) *massive multiple-input multiple-output* (MIMO) são tecnologias promissoras para as próximas gerações de redes de banda larga móvel. Nesses sistemas, o equipamento do usuário (UE - *user equipment*) é associado a um subconjunto de pontos de acesso (APs - *access points*) distribuídos na área de cobertura, resultando em melhorias na macrodiversidade e na eficiência espectral (SE - *spectral efficiency*) da rede, quando comparado com sistemas celulares convencionais. Apesar dos benefícios provenientes destes sistemas, desafios como métodos escaláveis de seleção de APs, complexidade computacional (CC) e intercoordenação entre unidades centrais de processamento (CPU - *central processing units*) ainda podem existir nos mesmos. Neste sentido, esta tese propõe um novo framework de seleção de APs que fornece escalabilidade a sistemas UC, possibilitando um uso mais eficiente dos recursos da rede, tais como potência de transmissão e demandas de processamento. A solução é baseada em uma decisão casada que visa estabelecer as conexões que sejam mais vantajosas tanto para os UEs quanto para os APs. Além disso, são propostas três estratégias que modificam os grupos de APs dos UEs, com o objetivo de reduzir o número de APs conectados a cada UE sem comprometer a SE. Resultados de simulação revelam que o framework de decisão casada pode melhorar a SE dos 95% likely UEs em até 163% quando comparado com as soluções de referência. Uma abordagem heurística que reduz os efeitos da inter coordenação entre CPUs também é proposta. A mesma diminui o número de UEs inter coordenados (i.e., UEs conectados a múltiplas CPUs) em cada CPU para reduzir as demandas de sinalização nos links de backhaul. Resultados numéricos indicam que o método proposto mitiga os impactos da inter coordenação entre CPUs, enquanto gera perdas marginais na SE e melhora a eficiência energética (EE). Por fim, investiga-se o desempenho de sistemas UC com capacidade de processamento limitada. Especificamente, assume-se que a CC de realizar a estimativa de canal e precodificação de sinais não cresce com o número de APs. Assim, o UE só pode ser associado a um número finito de APs. Além disso, propõe-se um método para ajustar os grupos de APs de acordo com a implementação da rede, ou seja, centralizada ou distribuída. Os resultados mostram que os sistemas UC podem manter a SE sob pequena degradação mesmo reduzindo a CC em até 96%. Além disso, o método proposto para ajustar o grupo de APs leva à reduções adicionais na CC.

Palavras-chaves: Abordagem centrada no usuário, *cell-free massive* MIMO, complexidade computacional, escalabilidade, inter coordenação entre CPUs, selecão de AP.

Abstract

User-centric (UC) cell-free (CF) massive multiple-input multiple-output (MIMO) systems are promising technologies for beyond 5G (B5G) networks. In these systems, the user equipment (UE) is associated with a subset of access points (APs) distributed into the coverage area, leading to improvements in macro-diversity and spectral efficiency (SE) compared to conventional cell-based systems. Despite the benefits, challenges such as scalable AP selection strategies, computational complexity (CC), and inter-central processing unit (CPU) coordination may still exist in these systems. In this regard, this thesis proposes a novel and general AP selection framework that affords scalability for UC systems, enabling more efficient use of the network resources, such as transmission power and reduced processing demands. The solution is based on a matched-decision among the most suitable connections for APs and UEs. Moreover, three strategies to fine-tune the AP clusters of UEs are proposed, aiming to reduce the number of APs connected to each UE without compromising the SE. Simulation results reveal that the matched-decision framework improves up to 163% the SE of the 95% likely UEs compared with baseline schemes. A heuristic approach that reduces the effects of inter-CPU coordination is also proposed. It decreases the number of inter-coordinated UEs (i.e., UEs connected to multiple CPUs) on each CPU to reduce signaling demands on backhaul links. Numerical results indicate that the proposed method mitigates inter-CPU coordination while yielding slight degradation in SE and improving energy efficiency (EE). Finally, this thesis investigates the performance of UC systems with limited processing capacity. Specifically, it is assumed that the CC of performing channel estimation and precoding signals does not increase with the number of APs. Thus, the UE can only be associated with a finite number of APs. Furthermore, a method is proposed for adjusting the AP clusters according to the network implementation, i.e., centralized or distributed. The results show that UC systems can keep the SE under minor degradation even if the CC up to 96%. Besides, the proposed method for adjusting the AP cluster leads to further reductions in CC.

Keywords: AP selection, computational complexity, cell-free massive MIMO, inter-CPU coordination, scalability, user-centric approach.

List of Figures

List of Tables

List of Abbreviations and Acronyms

4G	Fourth-Generation of Wireless Networks
5G	Fifth-Generation of Wireless Networks
AP	Access Point
ASD	Angular Standard Deviation
BS	Base Station
B5G	Beyond 5G
CSI	Channel State Information
CF	Cell-Free
CDF	Cumulative Distribution Function
CPU	Central Processing Unit
CoMP-JT	Coordinated Multi-Point with Joint Transmission
CDL	Clustered Delay Line
CC	Computational Complexity
DL	Downlink
DCC	Dynamic Cooperation Clustering
DAS	Distributed Antenna Systems
D-MIMO	Distributed MIMO
D-mMIMO	Distributed Massive MIMO
EE	Energy Efficiency
FDD	Frequency-Division Duplexing
HCPP	Hard Core Point Process
i.i.d	Independent and Identically Distributed
IC	Inter-Coordinated
LP-MMSE	Local Partial Minimum Mean-Squared-Error

L-MMSE	Local MMSE
LSFB	Largest-Large-Scale-Fading-Based
LoS	Line-of-Sight
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean-Squared Error
MR	Maximum-Ratio
MD	Matched-Decision
MSE	Mean Square Error
NS	Non-Scalable
NF	Noise Figure
NLoS	Non-Line-of-Sight-Wireless
P-MMSE	Partial Minimum Mean-Squared-Error
P-RZF	Partial Regularized Zero-Forcing
QuaDRiGa	Quasi Deterministic Radio Channel Generator
QoE	Quality of Experience
RV	Random Variable
RSRP	Reference Signal Received Power
RF	Radio-Frequency
SE	Spectral Efficiency
SINR	Signal-to-Interference-Plus-Noise Ratio
STD	Standard Deviation
SCF	Scalable Cell-Free
TDD	Time-Division Duplexing
TR	Technical Report
TDL	Tapped Delay Line
UL	Uplink

UE	User Equipment
UCC	User-Centric Clustering
UMi	Urban Micro
UC	User-Centric
ULA	Uniform Linear Array

List of Symbols

τ_c	Number of complex valued samples in each coherence block
τ_p	Number of complex valued samples dedicated for UL pilot signals
τ_u	Number of complex valued samples dedicated for UL data signals
τ_u	Number of complex valued samples dedicated for DL data signals
τ_s	Total number of coherence blocks
λ	Transmission wavelength
v	UE speed
$\mathbf{\Phi}$	Pilot book matrix
ψ_{kl}	Random phase shifts in LoS components
κ_{kl}	Rician factor
p_{LoS}	Probability of LoS component's existence
φ_{kl}	Azimuth angle
θ_{kl}	Elevation angle of the LoS component
β_{kl}	Large-scale fading gain
$\boldsymbol{\phi}_t$	Pilot vector t
η_i	Power that the UE i transmits in the UL direction
ϱ_k	Power that each AP serving the UE k assigns to it in centralized implementation.
ϱ_{kl}	Power that each AP l serving the UE k assigns to it in distributed implementation.
ϱ_d	Total transmission power of each AP
$\mathbf{\Psi}_{t_k l}$	Correlation matrix of UEs sharing the pilot t_k
ξ'	Regularization factor to estimate the UE k correlation matrix
ξ''	Regularization factor to estimate the correlation matrix of all UEs sharing the pilot t_k

υ'	Large-scale fading exponent for the DL fractional power allocation
κ'	Precoding vector exponent for the DL fractional power allocation
α	Pre-log factor term associated with estimating correlation matrices
ν_l	Efficiency of the power amplifier in the AP l
γ	Threshold gain in dB
$\delta\%$	Subset of APs that contribute most to the sum of the UE's total channel gain, in percentage
$\Gamma\%$	Subset of UEs that receive most of the power allocated by the AP
ε	The maximum loss limit allowed for the SE.
ζ	Project parameter to fine-tune the AP clusters based on the EE
λ_l	Partial channel strength indicator in the distributed implementation
λ_k	Partial channel strength indicator in the centralized implementation
Θ	Threshold to avoid excessive AP cluster adjustments

Notations

$\mathbf{V}$	Boldface uppercase letters denote matrices
$\mathbf{v}$	Boldface lowercase letters denote column vectors
$(\,.\,)^{\mathrm{T}}$	Transpose operation
$(\,.\,)^{\mathrm{H}}$	Conjugate-transpose operator
$\mathbb{E}\{\,.\,\}$	Expectation operator
$\lvert\mathcal{A}\rvert$	Cardinality of the set $\mathcal{A}$
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathrm{diag}(\mathbf{A}_n)$	Block-diagonal matrix with square matrices on the diagonal
$\lVert\,.\,\rVert$	Euclidean norm
$\mathrm{tr}(\,.\,)$	Trace of a matrix
$\mathcal{N}_{\mathbb{C}}\left(\mu, \sigma^2\right)$	Complex Gaussian RV, where μ is the mean and σ^2 is the variance
$\wedge$	Logical operation AND
$\vee$	Logical operation OR
$\mathcal{O}$	Big O

Contents

1 Introduction

1.1 Contextualization

The rapid expansion in the number of connected devices and the increasing popularization of high-capacity applications, such as real-time videos and streaming, has driven the demand for increasingly higher data rates on mobile broadband networks [1, 2]. According to Ericsson's mobility report, the total mobile data traffic is expected to increase threefold between 2023 and 2029 due to factors such as improved device capabilities and data-intensive content [3]. Therefore, wireless networks will have to provide greater spectral efficiencies (SEs) and a higher quality of experience (QoE) for connected user equipments (UEs) in the upcoming years.

In order to address these challenges, massive multiple-input multiple-output (MIMO) technology has emerged as a promising solution. Massive MIMO systems employ many antennas for signal transmission and reception, enabling UEs to utilize the same time-frequency resources through spatial multiplexing. As a result, the network's SE improves without increasing the transmission power [4].

Initially, the massive MIMO concept was applied to cellular networks and demonstrated great potential in providing increasingly higher SEs, making cellular massive MIMO a key technology for fifth-generation of wireless networks (5G) [1]. However, inter-cell interference and signal-to-interference-plus-noise ratio (SINR) degradation at cell edges can limit the performance of cellular massive MIMO systems [5–9]. In other words, even though massive MIMO is a promising technology, cellular networks have their physical limitations, which prevent the full potential of massive MIMO from being explored. Consequently, some requirements of beyond 5G (B5G) networks may not be met by cellular massive MIMO systems, e.g., uniform coverage and a large number of connected devices [10–15].

Distributed MIMO (D-MIMO) systems were a possible solution investigated in the literature to deal with cellular network coverage issues. In these ones, the coverage area is divided into disjoint static access point (AP) clusters, where each cluster comprises a subset of APs and serves the UEs within its zone, as depicted in Fig. 1. Thus, instead of connecting to a single base station (BS), as in cellular networks, the UE connects to multiple APs. Hence, the coverage probability improves as the UE receives signals from APs located in different positions [15,16]. D-MIMO systems are recognized by many names in the literature, including network MIMO, distributed antenna systems (DAS), and coordinated multipoint with joint transmission (CoMP-JT) [17–19]. D-MIMO systems were also employed in fourth-generation of wireless networks (4G), but despite the theoretical potential, they only provided small practical gains. This was mainly due to the intense interference coming from adjacent AP clusters [20, 21].

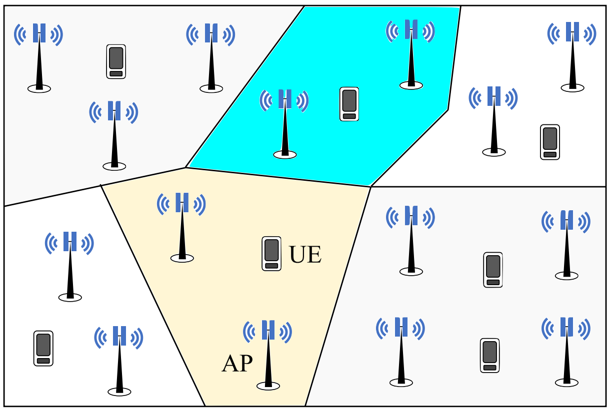


Figure 1 – Illustration of a D-MIMO network. The coverage area is divided into static and disjoint AP clusters. Each AP cluster provides service for the UEs within its zone.

In this context, user-centric (UC) cell-free (CF) massive MIMO networks have emerged to exploit the advantages of massive and distributed MIMO systems. In UC systems, several APs are spread out in the coverage area, and the UE is served by a subset of APs, called AP cluster, as Fig. 2 illustrates. Therefore, there are no cell boundaries from the UE's perspective during uplink (UL) and downlink (DL) transmission since all APs affecting a UE will actively participate in the communication. Furthermore, they mitigate network edge effects by moving the APs closer to the UEs, which means that although the UE is far from an AP, it can be close to another AP. Consequently, UC systems provide a more uniform SE and a better coverage probability than cell-based ones due to the enhanced macro-diversity and reduction of AP-UE distances [22–28].

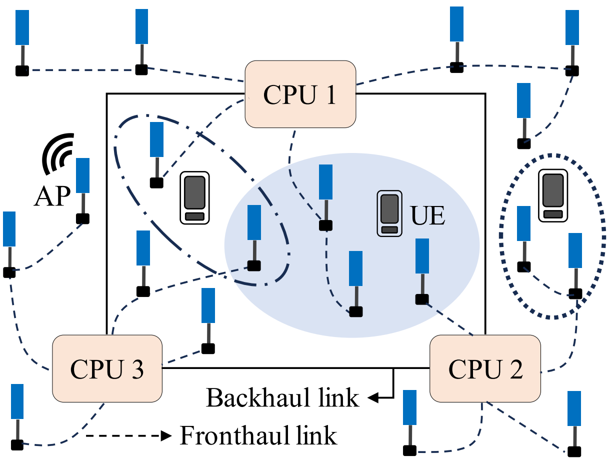


Figure 2 – Illustration of a UC CF massive MIMO network.

UC systems were initially called canonical CF massive MIMO, which assumed that all APs in the network served all UEs. However, this approach proved to be unfeasible due to the intense signaling required on the fronthaul links and the computational costs required from the central processing units (CPUs), which had to receive and process data from all UEs coming from all APs. Therefore, the UC approach also demand fewer fronthaul/backhaul requirements than its predecessor, the canonical CF massive MIMO, owing to the ability of each UE to connect to a subset of APs in its vicinity [24, 25].

UC systems can also achieve scalability when the network resources per AP, such as signal processing, signaling on fronthaul/backhaul, and total power, are independent of the number of UEs [27]. That is, the network resources per AP remain sufficient to provide service to the UEs even if the number of UEs goes to infinity. This would not be possible in a non-scalable system, such as the canonical CF, as the AP's resources (e.g., transmission power) would have to be divided among all UEs. Moreover, non-scalable systems may require enormous processing requirements of specific APs [26]. One can note that UC CF CF massive MIMO networks are a particular case of D-MIMO, i.e., a D-MIMO network with a single static cluster and a large number of APs. Hence, these networks can also be referred to as UC distributed massive MIMO (D-mMIMO) networks [29–32]. Besides, it is noteworthy that UC systems have been envisaged as one of the most promising technologies for B5G networks [33–36].

Even though the seminal papers embrace the UC approach, it can be impractical and unfair in aspects like processing capabilities, AP selection, and scalability. Moreover, computational complexity (CC) and signaling can still be a drawback in UC systems. In the next section, we highlight the challenges of UC systems investigated in this thesis.

1.2 Challenges in User-Centric Systems

UC CF massive MIMO systems are emerging technologies for the next generations of mobile communication networks. Thus, several challenges, such as channel estimation, pilot contamination, and power allocation, are still under investigation. Although some solutions have been proposed [37–51], there is still considerable exploration in these areas [52–59]. This section outlines the challenges investigated in this thesis, focusing on AP selection processes, processing capacity limitations, signaling demands, and scalability. For a deeper investigation into the other topics, one can refer to [37–59].

1.2.1 AP Processing Capabilities, Scalability and AP Selection

The distributed architecture employed by UC networks may require high processing capabilities from the APs since the APs need several advanced hardware components to process the signals of many UEs, such as a clock circuit and signal processor. Regarding scalability, the UC approach does not ensure that the network resources are independent of the number of UEs. That is, although traditional UC systems typically connect each UE to a subset of APs, most approaches do not present any mechanisms to prevent the network complexity from growing with the number of UEs. To achieve a scalable system, the complexity and resource requirements for each AP must remain finite when the number of UEs goes to infinity. In other words, the following tasks cannot rely on the number of UEs: (i) signal processing for channel estimation; (ii) signal processing for data reception and transmission; (iii) fronthaul signaling for data and channel state information (CSI) sharing; and (iv) power control optimization [26].

In a traditional UC system, these conditions cannot be met, as it does not limit the number of UEs an AP serves. For instance, if all UEs group near a specific AP, this AP will serve all these UEs, generating an immense amount of traffic on the fronthaul link that connects the AP to the CPU, when the number of UEs is large. In addition, the power and processing capabilities of the AP may not be sufficient to serve all UEs. In a nutshell, the complexity of implementing traditional UC systems can grow linearly or faster with the number of UEs, making the UC approach unscalable [60, 61]. A possible strategy to partially solve these drawbacks is to restrict the maximum number of UEs that each AP can serve [25, 26]. Such an approach can also alleviate the processing capabilities demands from APs. Therefore, scalable AP selection algorithms that provide scalability for baseline solutions are fundamental.

Regarding the AP selection process, most previous works do not limit the maximum number of UEs that each AP can serve, making them unscalable [26]. Additionally, many of these AP selection methods are unfair since they do not prevent the worst UEs from being dropped. To the best of our knowledge, previous works also do not consider a matched-decision among the APs and UEs, i.e., which connections among UEs and APs are more beneficial for both. They regularly assume that the UEs select a subset of APs to connect [24] or that the APs select a subset of UEs to serve [25]. However, relying on matched-decision AP selection is expected to improve the system performance significantly. Moreover, the AP clusters generated by AP selection methods can comprise APs that contribute only marginally to the UE's performance, leading the APs to waste power with UEs that do not take advantage of the allocated power. Therefore, strategies that fine-tune the AP clusters of UEs while keeping the SE under minor degradation or improving energy efficiency (EE) are also indeed necessary.

1.2.2 Computational Complexity and AP Cluster Adjustment

Several baseline solutions consider that the complexity of UC systems grows with the number of UEs and APs, which is not practical [23, 24]. In this regard, [26, 27] proposed a framework to provide scalability to UC systems. Essentially, it limits the number of UEs each AP can serve simultaneously. Consequently, the network resources (i.e., processing requirement, fronthaul/backhaul signaling, and total power) remain finite even if the number of UEs goes to infinity. The authors showed that scalable UC systems can still provide uniform coverage with negligible SE losses compared to the case when the UEs are served by all APs. The conclusions hold for both centralized and distributed network implementations. In the former, channel estimation and combining processing tasks are carried out on the CPUs, while in the latter, they occur on the APs. However, although the network resources become independent of the number of UEs, the signal processing complexity can still grow with the number of APs [26]. For instance, the number of complex multiplications required to perform channel estimation and precoding signals remains proportional to the number of APs serving the UE [27]. Thus, a more in-depth investigation into this topic is necessary, as the literature regularly assumes that there are more APs than UEs in the network.

Another limitation inherent to UC CF massive MIMO systems is that the AP selection processes are not adapted to the network implementations (i.e., centralized or distributed processing). They generally only intend to improve some key points, such as effective channel gain [62], reduce pilot contamination [26], among others [63–68]. Consequently, AP clusters may benefit one implementation over another. For instance, AP clusters with a large number of APs can degrade the EE and CC of UC systems operating in distributed implementation while they can improve the SE for the centralized ones.

1.2.3 Inter-CPU Coordination

Existing UC CF massive MIMO systems rely on a network composed of several APs linked by fronthaul to multiple CPUs, as Fig. 2 depicts [61]. Therefore, the AP cluster of each UE can comprise APs connected to different CPUs. Hence, the CPUs may need to exchange signals via backhaul to inter-coordinate several AP clusters, which means that the UC approach may require a lot of signaling and data sharing among the CPUs, i.e., intense inter-CPU coordination, also called inter-CPU communication [10]. The latter can limit coherent transmissions in wide-scale networks with more UEs and APs, impairing the feasibility of UC systems for large deployments. Thus, strategies to control the effects of inter-CPU coordination are crucial to improving system performance in terms of reducing signaling requirements. Nonetheless, such a task is challenging, as a threshold must be found between the degradation of SE and the reduction of inter-coordination effects between CPUs.

1.3 Related Works

The CF massive MIMO literature presents several AP selection schemes [62–68]. For instance, [62] associates the UE with the AP, simultaneously presenting the largest channel gain and causing less interference. In contrast, [65] utilizes a graph neural network-based AP selection to reduce the number of reference signal received power (RSRP) measurements necessary to generate the AP clusters. Nonetheless, the complexity of [62] increases with the number of UEs, and the complexity of [65] can grow faster with the number of APs compared to other baseline solutions such as [26].

These are some strategies that the literature has proposed, but the most commonly analyzed methods are presented in [23–26]. In [23], the AP cluster of each UE is composed of all APs, which is the canonical version of CF. This method improves the SE of the worst UEs and increases the network's coverage probability compared to co-located systems. However, it can require exceptionally high computation costs and backhaul/fronthaul capabilities from the APs. Moreover, all APs have to divide their transmission powers among all UEs, leading the power allocation to be as small as zero in some cases. Therefore, the canonical form of CF is non-scalable (NS).

In [25], the APs serve the UEs with the largest estimated channel in their vicinity. This scheme improves the SE of the UEs compared to canonical CF systems and avoids the depletion of allocated power by limiting the number of UEs each AP can serve. Nonetheless, it is an NS method since the APs need to have access to the channel estimates of all UEs in the network. Additionally, it does not prevent the worst UEs from being dropped. In [24], the UE establishes a connection with the subset of APs that contribute most to the sum of its total channel gain, which can improve network energy efficiency. Nevertheless, it is also an NS scheme since it does not limit the number of UEs each AP can serve. Therefore, it does not guarantee scalability for the network resources.

In [26], a scalable AP selection method that relates the pilot allocation to the AP cluster formation was proposed. The solution is based on the dynamic cooperation clustering (DCC) framework and prevents the worst UEs from being dropped. Moreover, the paper limits the number of UEs per AP and proves that only heuristic solutions that do not rely on the number of UEs are scalable. However, the literature in this research field is still in its infancy, and further investigations are indeed required to provide more valuable insights and advances in the area.

Regarding reducing the network complexity under computational and signaling aspects, the literature has also proposed some approaches [25–27]. For instance, [25] introduced the UC approach, demonstrating that UC systems could achieve comparable performance to canonical CF massive MIMO systems while reducing CC and fronthaul requirements. In [26, 27], the authors analyzed the scalability of UC systems, presenting their performance in terms of SE assuming both centralized and distributed network implementations. The authors demonstrated that the CC of the network and signaling in the fronthaul links could be prevented from growing with the number of UEs, but without providing any analysis regarding the number of APs. Besides, [26, 27] claimed that their proposed strategies are effective for UC systems with multiple CPUs but did not detail the network's requirements to make it successful. That is, the authors in [26, 27] did not investigate the impacts of inter-CPU coordination.

Strategies for reducing the number of APs serving the UE were proposed in [63,69]. Nevertheless, a mechanism to prevent network processing demands from growing with the number of APs was not presented, i.e., the maximum number of APs serving each UE was not restricted. In [52], the maximum number of APs serving the UE was limited, defined as a parameter that can be adjusted to avoid losses in SE. However, the analysis did not account for the system processing capacity limitation. In addition, a detailed investigation regarding CC and multiple CPUs was not provided.

In [10], an approach to mitigate inter-CPU communication was proposed. The authors considered a network composed of multiple virtual cells, each managed by an individual CPU. The UEs within a virtual cell are exclusively associated with the APs inside that cell. Conversely, UEs at the cell edges can connect to APs from different virtual cells (i.e., belonging to distinct CPUs). This approach reduced the effect of inter-CPU communication compared to traditional UC systems. Despite this advantage, the SE can decrease, while the signaling demands between CPUs still grow with the number of UEs. Regarding the adjustment of AP clusters under different network implementations, to the best of the authors' knowledge, no other works addressing this topic were found.

1.4 Thesis Contributions

This section presents the main contributions of this thesis. It summarizes this thesis's proposals to fill the literature gaps regarding the drawbacks of UC CF massive MIMO systems mentioned above.

The first contribution of this thesis is a general and scalable AP selection framework that exploits a matched-decision among UEs and APs. The proposed method is divided into two stages, where the UEs first connects to an intermediate subset of APs and then to a final cluster of APs. The first stage allows the UEs and APs to establish the best connection for both in terms of large-scale fading. The second one enables the UEs to expand their AP clusters, aiming to improve SE. Additionally, modifying some parameters allows the proposed algorithm to behave like previous AP selection methods. Nevertheless, it improves these schemes by affording scalability and decreasing their time complexity. We compare the proposed AP selection strategy with the canonical CF system and three baseline solutions, one of which is also a scalable approach. We also analyze the system performance for perfect and imperfect knowledge of channel statistics.

Secondly, three novel methods for fine-tuning the AP clusters are proposed, cleverly dropping UE-AP connections that do not significantly contribute to the system performance. To the best of the author's knowledge, this thesis contains the first work that proposes general strategies for fine-tuning the AP clusters of UEs in scalable UC systems. The first fine-tuning scheme relies on power allocation, the second is based on SE, and the third is on EE. The proposed techniques work for any AP selection scheme, and numerical results indicate that the SE can be kept under minor degradation, even when dropping the UEs' connections with some APs.

This thesis also investigates the performance of scalable UC CF massive MIMO systems by assuming that the CC to perform channel estimation and precoding signals does not grow with the number of APs. In particular, it is considered a UC system where the UE is associated only with a finite number of APs, i.e., the UE is connected only with the APs having the strongest channel gains. In other words, the maximum AP cluster size of the UE is controlled.

To the best of the author's knowledge, this thesis presents a pioneering work that proposes an approach to constrain the CC of UC systems from growing with the number of APs. Moreover, a method is proposed to adjust the AP clusters according to the network implementation. The proposed methods for adjusting the AP clusters work in UC systems with and without processing capacity limitations, and can be used as an alternative solution for reducing CC in UC systems without processing capacity limitations. As far as the author is aware, this thesis is also innovative in proposing a method for adjusting the AP clusters according to the network implementations.

Finally, this thesis proposes a novel method that reduces the effects of inter-CPU coordination in UC CF massive MIMO systems while keeping the SE under minor degradation. The proposed method assumes that each CPU serves only a finite number of inter-coordinated UEs, i.e., UEs connected to more than one CPU. Hence, the CPUs can drop inter-coordinated UEs with the smallest channel gains to mitigate inter-CPU coordination. To the best of the author's knowledge, the proposed approach is the first method that effectively limits the impacts of inter-CPU coordination in UC CF massive MIMO systems.

1.5 Objectives

The main objective of this thesis is to enhance the performance of scalable UC CF massive MIMO systems. This is achieved through the proposal of AP selection strategies, fine-tuning algorithms, and AP cluster adjustment, aiming to improve SE and EE, reduce CC, and decrease the impacts of inter-CPU coordination.

The following specific objectives are addressed to achieve the main objective:

- Investigate the impact of the matched-decision scheme in UC systems considering different processing capacity restrictions in the APs, i.e., by varying the number of UEs each AP can serve.

- Compare the performance of the matched-decision framework with baseline AP selection schemes and afford scalability to non-scalable baseline solutions. Perform the comparisons for perfect and imperfect knowledge of channel statistics.
- Verify the effectiveness of the proposed fine-tuning algorithms in reducing the number of APs serving the UEs while keeping the SE under minor degradation or improving the EE of UC systems.
- Limit the CC of performing channel estimation and generating the precoding vectors. Then, verify how much it affects the network's SE. Moreover, assess the pros and cons of adjusting the AP clusters according to each network implementation.
- Investigate the impacts of reducing inter-CPU coordination in UC systems by restricting the number of inter-coordinated UEs each CPU can serve. Perform analyses regarding the SE, EE, and number of inter-coordinated UEs per CPU.
- Consider a wide range of UEs, APs, and antennas per AP in the simulation results. Assume that each AP can serve a limited number of UEs, where this number varies according to the scenario. Evaluate the system performance under different degrees of cooperation among the APs, i.e., centralized and distributed network implementations.

1.6 Thesis Organization

The remainder of this thesis is organized as follows:

Chapter 2 - User-Centric Cell-Free Massive MIMO Networks

It describes the system model utilized in this thesis to simulate a UC CF massive MIMO network. The channel estimation procedure, the mathematical representation of AP clusters, the network implementations, power allocation, and the DL SE are presented. The model adopted for EE and the motivation behind the CF concept are also presented.

Chapter 3 - Matched-Decision AP Selection

The chapter presents the framework proposed in this thesis to enable a matched-decision among the most suitable connections to UEs and APs in UC systems. Moreover, the chapter outlines the three methods proposed in this thesis to fine-tune the AP clusters of UEs. The algorithms and equations utilized in the proposed strategies are presented, and their time complexity is discussed.

Chapter 4 - Reducing Inter-CPU Coordination

This chapter describes the proposed strategy to reduce the effects of inter-CPU coordination in UC CF massive MIMO systems. It details the procedure for associating the UE with a primary CPU and shows how CPUs can drop UEs to mitigate the effects of inter-CPU coordination.

Chapter 5 - Scalable User-Centric Cell-Free Massive MIMO with Limited Processing Capacity

It discusses the proposed approaches to reduce the CC of UC systems in performing channel estimation and generating the combining vectors. It details the proposed procedure and shows that it is possible to prevent the CC from growing with the number of APs by controlling the maximum AP cluster size of each UE. In addition, the two methods that adjust the AP clusters according to the network implementations are presented.

Chapter 6 - Numerical Results

This chapter presents illustrative numerical results along with insightful discussions to demonstrates the effectiveness of the proposed methods compared to previous baseline schemes. For instance, it shows that the matched-decision scheme can improve up to 163% the SE of the 95% likely UEs compared with baseline solutions and that the number of UEs that each AP serves is crucial for improving SE. Simulation results also reveal that UC systems can maintain the SE under small degradation even if the processing capacity is restricted, decreasing the CC by up to 96%. Additionally, they show that it is possible to control the effects of inter-CPU coordination without significantly harming the SE.

Chapter 7 - Conclusions and Future Works

This chapter concludes the thesis, presenting final considerations and possibilities for future works.

2 User-Centric Cell-Free Massive MIMO Networks

This chapter presents the theoretical foundation of this thesis associated with UC CF massive MIMO system. The system model is presented, introducing channel models, estimation methods, precoding schemes, and others. The assumptions adopted in this thesis to compute performance metrics such as SE and EE are also discussed.

2.1 System Model

We consider a CF massive MIMO network composed of K ingle-antenna UEs, J CPUs, and L APs distributed over the coverage area, where $L > K$. Each AP is equipped with N antennas and the total number of antennas considering all APs is $M = NL$. The APs connect to the CPUs through fronthaul links, while the CPUs communicate with each other through backhaul ones, as depicted in Fig. 2. The network infrastructure (e.g., fronthaul, backhaul, and core network) is assumed to be perfectly syncronized, error-free and capable of supporting data traffic, which is a common assumption in the CF literature [22–28]. Other approaches may also consider fronthaul/backhaul limitations when modeling a UC CF massive MIMO network [70–76]. However, we focus on the most common assumptions to compare the proposed methods and baseline solutions fairly.

2.1.1 Block Fading Model and TDD Protocol

In wireless communication systems, the propagation channels are generally time-variant. Consequently, their frequency response can present fluctuations over time. These oscillations, known as fading, represent variations in signal strength experienced by radio waves as they travel through the wireless channel, often caused by factors such as multipath propagation and environmental changes like mobility and obstacles. Even slight variations in the propagation environment, such as movements of the receiver on the scale of centimeters or millimeters, can significantly change the channel. Nevertheless, if we consider a time interval sufficiently short, the frequency response becomes approximately constant and flat-fading. This time interval is called coherence time, and the frequency range over which the channel frequency response is flat-fading is named coherence bandwidth.

A coherence block corresponds to the time-frequency block, containing both the coherence time and bandwidth. Each coherence block undergoes an independent channel realization, requiring channel properties to be estimated separately for each coherence block. Besides, since it is constant, the channel frequency response can be represented by only one scalar coefficient in each coherence block.

UC CF massive MIMO systems typically utilize the block fading model to describe fading channels. In this one, the time-frequency plane is divided into several coherence blocks, as illustrated in Fig. 3. Moreover, each coherence block is assumed to operate in time-division duplexing (TDD) mode, which is a transmission protocol where the UL and DL channels occur at the same frequency band but are separated in time. Therefore, the UL and DL channels utilize the same coherence bandwidth, but occur at different time intervals within each coherence block. That is, the time interval of each coherence time is divided between DL and UL signals.

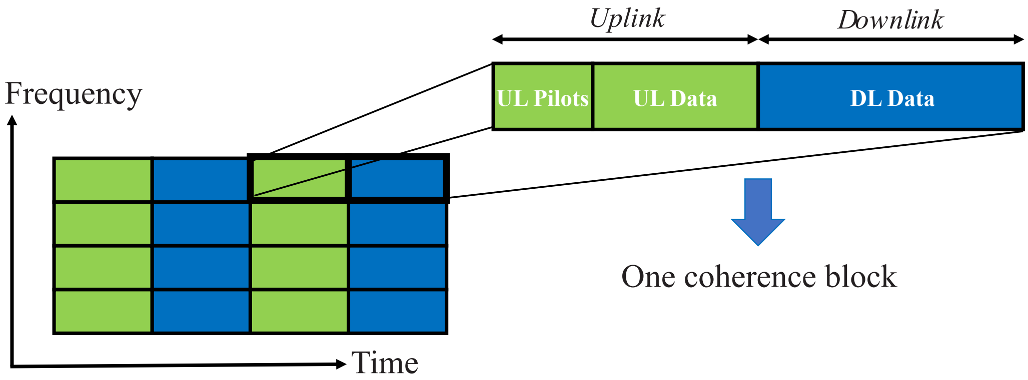


Figure 3 – The block fading model divides the time-frequency plane into several coherence blocks, where the channel is frequency-flat and time-invariant. In each coherence block, the UL and DL channels use the same coherence bandwidth, but occur at distinct time intervals within each coherence time.

Source: adapted from [4].

The coherence block is calculated as $\tau_\mathrm{c} = T_\mathrm{c} B_\mathrm{c}$, where T_c and B_c denote the coherence time and bandwidth, respectively. The coherence block is measured in complex-valued samples and represents the symbols utilized to convey information in a UC system. In practice, the exact values of T_c and B_c are challenging to model since they can be a function of several factors, such as the UE mobility, carrier frequency, and delay spread. The latter is the time interval between the shortest and longest paths the transmitted signal takes before reaching the receiver. A common approximation computes $T_\mathrm{c} = \lambda/4v$ and $B_\mathrm{c} = 1/2T_\mathrm{d}$, where λ denotes the transmission wavelength, v means the UE speed, and T_d is the delay spread [4].

In TDD mode, the complex-valued samples τ_c of each coherence block can be divided into τ_p samples for UL pilot signaling, and τ_u and τ_d samples for UL and DL data transmissions, respectively, as Fig. 4 shows. The pilot signals are predefined sequences of τ_p-length that the UEs periodically send to the APs to estimate their channels. Then, the APs (or CPUs) correlate the received pilot with the predefined pilot signal and estimate the UE channel. The pilot signals are mutually orthogonal to each other, with each pilot being one of the columns of matrix $\boldsymbol{\Phi} \in \mathbb{C}^{\tau_p \times \tau_p}$, referred to as pilot book [4].

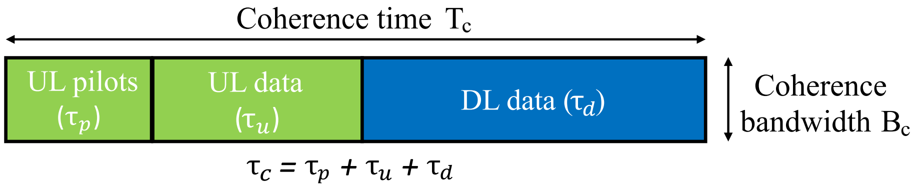


Figure 4 – A TDD protocol is considered where each coherence block can be divided into pilot signaling, UL data, and DL data.

The pilot signals are sent only in the UL direction since the channel frequency response remains constant throughout T_c, making UL channel estimates also valid for DL transmissions. Hence, massive MIMO systems (including UC CF massive MIMO) usually consider that the UL and DL channels are reciprocal in the coherence time T_c. That is, the UL and DL estimated channels are equal since they utilize the same coherence bandwidth in TDD mode. In this thesis, we focus on DL transmissions and consider that $\tau_u = 0$.

The principle of reciprocity is one of the characteristics of wireless communication channels that the TDD protocol benefits from in massive MIMO systems, as the reciprocity reduces the CSI acquisition overhead [77]. However, there are many efforts to find efficient implementations based on frequency-division duplexing (FDD) since a vast amount of spectrum is reserved for FDD operation [78–84]. Furthermore, DL channel estimation can also be helpful for decoding data from UEs, as reciprocity is not always achieved in practice, as hardware imperfections, impedance mismatches, and calibration errors in the radio-frequency (RF) chain can make the channel non-reciprocal [85, 86].

2.1.2 Channel Models

It is considered that the channel vector $\mathbf{h}_{kl} \in \mathbb{C}^{N\times 1}$ between the AP l and UE k undergoes an independent correlated Rician fading, being defined as [87–89]

$$\mathbf{h}_{kl} = \underbrace{\sqrt{\frac{\kappa_{kl}}{1+\kappa_{kl}}}\mathbf{h}_{kl}^{\mathrm{LoS}}\, e^{j\psi_{kl}}}_{\overline{\mathbf{h}}_{kl}} + \underbrace{\sqrt{\frac{1}{1+\kappa_{kl}}}\mathbf{h}_{kl}^{\mathrm{NLoS}}}_{\widetilde{\mathbf{h}}_{kl}}, \tag{2.1}$$

where $\overline{\mathbf{h}}_{kl} \in \mathbb{C}^{N\times 1}$ means the deterministic line-of-sight (LoS) component, while $\widetilde{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \widetilde{\mathbf{R}}_{kl}) \in \mathbb{C}^{N\times 1}$ stands for the small-scale fading with statistical correlation matrix $\widetilde{\mathbf{R}}_{kl} = \mathbb{E}\{\widetilde{\mathbf{h}}_{kl}\widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} \in \mathbb{C}^{N\times N}$. The small-scale fading denotes the effects of the non-line-of-sight-wireless (NLoS) propagation. The correlation matrix represents the large-scale fading of the system, being a function of the spatial channel correlation, path loss, and shadowing. The term $\psi_{kl} \sim \mathcal{U}[0, 2\pi)$ denotes random phase shifts that may occur in LoS components due to the UE's mobility, and the Rician factor κ_{kl} stands for the power ratio between the

LoS and NLoS components, being defined as [90]

$$\kappa_{kl} = \frac{p_{\mathrm{LoS}}}{1 - p_{\mathrm{LoS}}}, \tag{2.2}$$

with p_{LoS} being the probability of the LoS component's existence. Thus, $p_{\mathrm{LoS}} = 0$ for propagation links presenting only NLoS components.

Assuming the APs are equipped with half-wavelength-spaced uniform linear arrays (ULAs), the correlation matrix of the NLoS channel $\mathbf{h}_{kl}^{\mathrm{NLOS}}$, i.e., $\mathbf{R}_{kl}^{\mathrm{NLoS}}$, can be computed following the local scattering model for spatial correlation presented in [4, Sec. 2.6]. Thus, the correlation matrix of the term $\widetilde{\mathbf{h}}_{kl}$ in (2.1) is given by

$$\widetilde{\mathbf{R}}_{kl} = \mathbb{E}\{\widetilde{\mathbf{h}}_{kl}\widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} = \frac{1}{\kappa_{kl} + 1}\mathbf{R}_{kl}^{\mathrm{NLoS}}, \tag{2.3}$$

which implies that $\mathbf{R}_{kl} = \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\mathrm{H}}\} = (\overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\mathrm{H}} + \widetilde{\mathbf{R}}_{kl})$. Moreover, the LoS channel between the UE k and AP l can be expressed as

$$\mathbf{h}_{kl}^{\mathrm{LoS}} = \sqrt{\beta_{kl}}\left[1, e^{\mathrm{j}\pi\sin(\varphi_{kl})\cos(\theta_{kl})}, \cdots, e^{\mathrm{j}(N-1)\pi\sin(\varphi_{kl})\cos(\theta_{kl})}\right]^{\mathrm{T}}, \tag{2.4}$$

where φ_{kl} denotes the azimuth angle, θ_{kl} is the elevation angle of the LoS component, and β_{kl} is the large-scale fading gain, which can be calculated as $\beta_{kl} = \mathrm{tr}(\mathbf{R}_{kl})/N$. We assume that the correlation matrices remain unchanged for many coherence blocks, while the small-scale fading changes in each block [11]. Additionally, the channels of different APs are assumed to be uncorrelated, thus $\mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl'}^{\mathrm{H}}\} = \mathbf{0}_N$ for $l \neq l'$ [91–94].

It is worth noting that there are different models for the channel between an AP and UE, which are valid under different assumptions. The Rician fading is often used in UC systems because it allows the APs to be closer to the UEs. Moreover, Rician fading has the advantage of capturing the effects of the LoS and NLoS components. Rayleigh fading propagation is also commonly used to model the channel vectors of UC systems. Nonetheless, it assumes that the LoS channel is blocked and relies on rich scattering scenarios, i.e., when there are many NLoS propagation components. An independent correlated Rayleigh fading realization can be expressed as

$$\mathbf{h}_{kl} = \sqrt{\mathbf{R}_{kl}^{\mathrm{NLoS}}}\,\mathbf{g}_{kl}, \tag{2.5}$$

where $\mathbf{g}_{kl} \in \mathbb{C}^{N\times 1}$ is composed of elements that are independent and identically distributed (i.i.d) complex Gaussian $\mathcal{N}_{\mathbb{C}}(0,1)$ random variables (RVs). One can note that (2.5) equals (2.1) for $p_{\mathrm{LoS}} = 0$. It is noteworthy that there are also other solutions for modeling the channel propagation, such as tapped delay line (TDL), clustered delay line (CDL), and quasi deterministic radio channel generator (QuaDRiGa) models. Nevertheless, a more profound discussion regarding the best channel model is outside the scope of this thesis.

2.1.3 Uplink Training

When a UE enters the network, the APs or CPUs need to know its channel to serve it. Thus, UC systems incorporate a training phase where the channels between UEs and APs are estimated. During the training phase, the UEs send pilot signals of τ_p-length to the APs to estimate their channels [4, 95]. Recall that in TDD mode, τ_p samples of each coherence block are dedicated to the pilot signals, which are vectors of dimension τ_p mutually orthogonal to each other.

Ideally, a unique orthogonal pilot signal should be assigned to each UE, so that $\tau_p = K$. However, coherence blocks have limited resources, making it impossible to assign a distinct pilot for each UE if the number of UEs K is greater than the number of complex-valued samples in the coherence block, i.e., $K > \tau_c$. Because of this, a pilot t_k can be reused by some UEs if $K > \tau_p$ to ensure scalability for the pilot resources, with $\tau_p < \tau_c$. The pilot reuse leads to the phenomenon known as pilot contamination, where channel estimation is less accurate, as the estimated channels of the UEs utilizing the same pilot correlate [4]. Therefore, choosing the most appropriate number of complex-valued samples for pilot signals τ_p is not a trivial task. A balance is generally sought between SE and estimation accuracy [32].

Given that the pilot vector $\boldsymbol{\phi}_t \in \mathbb{C}^{\tau_p \times 1}$ must satisfy $|\boldsymbol{\phi}_t|^2 = \tau_p$, for $t = 1, .., \tau_p$, and that $\boldsymbol{\phi}_{t_1}^{\mathrm{H}} \boldsymbol{\phi}_{t_2} = 0$, for $t_1 \neq t_2$, the received signal $\mathbf{Y}_l^{\text{pilot}} \in \mathbb{C}^{N \times \tau_p}$ at AP l during the channel estimation phase is

$$\mathbf{Y}_l^{\text{pilot}} = \sum_{i=1}^{K} \sqrt{\eta_i} \mathbf{h}_{il} \boldsymbol{\phi}_{t_i}^{\mathrm{T}} + \mathbf{N}_l \tag{2.6}$$

where η_i denotes the power that the UE i transmits in the UL direction and $\mathbf{N}_l \in \mathbb{C}^{N \times \tau_p}$ represents the noise at the receiver with i.i.d elements distributed as $\mathcal{N}_{\mathbb{C}}(0, \sigma_{\text{ul}}^2)$. Considering that the channel $\mathbf{h}_{kl}$ needs to be estimated and that the UE k transmitted the pilot $t_k = t_1$, the expression of the received signal can be further simplified. To do this, $\mathbf{Y}_l^{\text{pilot}}$ is multiplied by the normalized conjugate of $\boldsymbol{\phi}_{t_k}$ (i.e, $\boldsymbol{\phi}_{t_k}^* / \sqrt{\tau_p}$) to eliminate the interference coming from UEs using pilots signals orthogonal to t_k. After this, only the effects of UEs sharing the pilot t_k will affect the received signal. Thus, letting $\mathcal{S}_{t_k} \subset \{1, ..., K\}$ represent the subset of UEs that send the pilot t_k to the APs, the received signal at AP l is given by

$$\mathbf{y}_{t_k l}^{\text{pilot}} = \mathbf{Y}_l^{\text{pilot}} \frac{\boldsymbol{\phi}_{t_k}^*}{\sqrt{\tau_p}} = \sum_{i=1}^{K} \frac{\sqrt{\eta_i}}{\sqrt{\tau_p}} \mathbf{h}_{il} \boldsymbol{\phi}_{t_i}^{\mathrm{T}} \boldsymbol{\phi}_{t_k}^* + \frac{1}{\sqrt{\tau_p}} \mathbf{N}_l \boldsymbol{\phi}_{t_k}^*, \tag{2.7}$$

which can be simplified to

$$\mathbf{y}_{t_k l}^{\text{pilot}} = \sum_{i \in \mathcal{S}_{t_k}} \sqrt{\tau_p \eta_i}\, \mathbf{h}_{il} + \mathbf{n}_{t_k l}, \tag{2.8}$$

where $\boldsymbol{\phi}_{t_k}^* / \sqrt{\tau_p}$ is a vector with unit norm and $\mathbf{n}_{t_k l} = \frac{1}{\sqrt{\tau_p}} \mathbf{N}_l \boldsymbol{\phi}_{t_k}^* \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma_{\text{ul}}^2 \mathbf{I}_N)$. Assuming that the correlation matrix $\mathbf{R}_{kl}$ is perfectly known and utilizing the linear minimum mean-squared error (MMSE) estimator, the estimated channel $\widehat{\mathbf{h}}_{kl}$ is given by

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p \eta_k}\, \mathbf{R}_{kl} \mathbf{\Psi}_{t_k l}^{-1} \mathbf{y}_{t_k l}^{\text{pilot}}, \tag{2.9}$$

where $\mathbf{\Psi}_{t_k l} \in \mathbb{C}^{N \times N}$ is a matrix containing the sum of the correlation matrices of UEs that share the pilot t_k, which can degrade the system performance since it causes pilot contamination. The computation of $\mathbf{\Psi}_{t_k l}$ can be performed as

$$\mathbf{\Psi}_{t_k l} = \mathbb{E}\{(\mathbf{y}_{t_k l}^{\text{pilot}})(\mathbf{y}_{t_k l}^{\text{pilot}})^{\text{H}}\} = \sum_{i \in \mathcal{S}_{t_k}} \eta_i \tau_p (\overline{\mathbf{h}}_{il} \overline{\mathbf{h}}_{il}^{\text{H}} + \widetilde{\mathbf{R}}_{il}) + \sigma_{\text{ul}}^2 \mathbf{I}_N. \tag{2.10}$$

Depending on the network implementation, UC systems can delegate the channel estimation process to APs or CPUs. These network implementations will be discussed further in Section 2.2. In addition, note that the estimated channel $\widehat{\mathbf{h}}_{kl}$ relies on the perfect knowledge of $\mathbf{R}_{kl}$ and $\mathbf{\Psi}_{t_k l}$ in (2.9). Nonetheless, channel statistics change in practice due to UE mobility or scheduling. Hence, each AP (or a subset of CPUs) needs to estimate these matrices, leading $\widehat{\mathbf{h}}_{kl}$ to

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p \eta_k}\, \widehat{\mathbf{R}}_{kl} \widehat{\mathbf{\Psi}}_{t_k l}^{-1} \mathbf{y}_{t_k l}^{\text{pilot}}, \tag{2.11}$$

where $\widehat{\mathbf{R}}_{kl}$ and $\widehat{\mathbf{\Psi}}_{t_k l}$ are the imperfect versions of $\mathbf{R}_{kl}$ and $\mathbf{\Psi}_{t_k l}$. We rely on [92] to calculate these terms, which proposes a two-stage approach. It considers that the AP l (or the CPUs) observes many channel realizations in different coherence blocks and uses the received pilots to estimate

$$\widehat{\mathbf{\Psi}}_{t_k l}^{(\text{sample})} = \frac{1}{N_\Psi} \sum_{n=1}^{N_\Psi} \mathbf{y}_{t_k l}^{\text{pilot}}[n] \left(\mathbf{y}_{t_k l}^{\text{pilot}}[n]\right)^{\text{H}}, \tag{2.12}$$

where $N_{\mathbf{\Psi}}$ is the number of observations. The AP computes $\widehat{\mathbf{\Psi}}_{t_k l}$ as $\widehat{\mathbf{\Psi}}_{t_k l} = \xi' \widehat{\mathbf{\Psi}}_{t_k l}^{(\text{sample})} + (1 - \xi') \widehat{\mathbf{\Psi}}_{t_k l}^{(\text{diagonal})}$, where $\xi' \in [0, 1]$ is a regularization factor and $\widehat{\mathbf{\Psi}}_{t_k l}^{(\text{diagonal})}$ is the main diagonal of $\widehat{\mathbf{\Psi}}_{t_k l}^{(\text{sample})}$. Assuming that the channel statistics are fixed over the system's bandwidth (B_s) and a time interval (T_s), the number of coherence blocks that the AP can observe is $N_{\mathbf{\Psi}} \leq \tau_s$, where $\tau_s = B_s T_s / \tau_c$. For instance, considering that each coherence block has $\tau_c = 200$ samples in a mobile scenario with $B_s = 100\,\text{MHz}$ and $T_s = 0.5\,\text{s}$, the AP could observe the channel over $\tau_s = 250000$ coherence blocks.

One can note that no extra-pilots are needed to calculate (2.12), since it is obtained from the pilots used to perform channel estimation. However, they are needed to compute $\widehat{\mathbf{R}}_{kl}$. To adjust the strategy proposed in [92] to a UC CF massive MIMO scenario, we consider that τ_p pilots are utilized to estimate $\widehat{\mathbf{R}}_{kl}$ over N_R observations, requiring a total of $N_R \tau_p K$ extra pilots[1]. $\widehat{\mathbf{R}}_{kl}$ is computed as $\widehat{\mathbf{R}}_{kl} = \xi'' \widehat{\mathbf{R}}_{kl}^{(\text{sample})} + (1 - \xi'') \widehat{\mathbf{R}}_{kl}^{(\text{diagonal})}$, where ξ'' is a regularization factor and $\widehat{\mathbf{R}}_{kl}^{(\text{sample})} = \widehat{\mathbf{\Psi}}_{t_k l}^{(\text{sample})} - \widehat{\mathbf{\Psi}}_{t_k l,-k}^{(\text{sample})}$. The term $\widehat{\mathbf{\Psi}}_{t_k l,-k}^{(\text{sample})}$ denotes a stage where only the interfering UEs sharing the same pilot as the UE k send the pilot.

[1] The scalability of [92] is questionable as its complexity increases with K. However, this thesis only investigates the influence of the imperfect knowledge of $\mathbf{R}_{kl}$ in the system performance. Therefore, a more profound discussion involving the best method of acquiring $\widehat{\mathbf{R}}_{kl}$ is out of the scope of this thesis. For a deeper investigation, one can read the following references [93, 94, 96].

The parameters ξ' and ξ'' can be optimized according to the evaluated scenario when the APs are equipped with more than one antenna ($N > 1$), especially for $N \gg 1$. For simplicity, we evaluate the impacts of $\widehat{\mathbf{R}}_{kl}$ and $\widehat{\mathbf{\Psi}}_{t_k l}$ only for single antenna APs ($N = 1$). Hence, $\xi' = 1$ and $\xi'' = 1$.

It is worth mentioning that channel estimation can be performed by alternative methods, such as element-wise MMSE and least square estimators. The former does not utilize the full spatial correlation matrices but only the main diagonal elements, while the latter neglects channel correlation. Both demand smaller CC than MMSE but at the cost of reduced estimation accuracy [4].

2.1.4 DL Data Transmissions

In UC systems, each UE is associated with a subset of APs called AP cluster. In order to generate AP clusters, we proceed as follows. First, let $\mathcal{M}_k \subset \{1, ..., L\}$ denote the indexes of APs that serve the UE k. Second, let $\mathbf{c}_k = [\mathrm{c}_{k1}, ..., \mathrm{c}_{kL}] \in \mathbb{N}^{1\times L}$ be the vector which designates the APs that transmit a signal to UE k. That is, if the AP l transmits a signal to UE k, $\mathrm{c}_{kl} = 1$, otherwise $\mathrm{c}_{kl} = 0$, which means that

$$\mathrm{c}_{kl} = \begin{cases} 1 & \text{if } l \in \mathcal{M}_k \\ 0 & \text{if } l \notin \mathcal{M}_k \end{cases}. \tag{2.13}$$

Moreover, the matrix $\mathbf{D}_{kl} \in \mathbb{N}^{N\times N}$ describes which antennas of the AP l establish a connection to the UE k. It is assumed that all N antennas of the AP l transmit a signal to the UE k. This is similar to write

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & \text{if } l \in \mathcal{M}_k \\ \mathbf{0}_N & \text{if } l \notin \mathcal{M}_k \end{cases}. \tag{2.14}$$

Note that $\mathbf{c}_k$ and the diagonal block matrix $\mathbf{D}_k = \mathrm{diag}\left(\mathbf{D}_{k1}, ..., \mathbf{D}_{kL}\right) \in \mathbb{C}^{M\times M}$ define the AP cluster of UE k, but this is not their only function. They also indicate the subset of UEs that each AP serves after the AP cluster formation. We denote this subset as $\mathcal{D}_l \subset \{1, ..., K\}$, where $\mathcal{D}_l$ has the indexes of the UEs that the AP l serves. Hence, it is possible to observe that $|\mathcal{D}_l| = \sum_{k\in\mathcal{D}_l} \mathrm{c}_{kl}$ and $|\mathcal{M}_k| = \sum_{l\in\mathcal{M}_k} \mathrm{c}_{kl}$. The cardinalities of $\mathcal{M}_k$ and $\mathcal{D}_l$ can also be represented by L_k and K_l, where $L_k = |\mathcal{M}_k|$ and $K_l = |\mathcal{D}_l|$.

The cardinality of $\mathcal{D}_l$ is regularly lower than K. However, if all UEs group in the vicinity of specific APs, K_l may be equal to K, which is unscalable. In order to solve this drawback, it is assumed that $K_l \leq U_{max}$, where U_{max} is a constant that remains unchanged even if $K \to \infty$. That is, even if the number of UEs is extremely high, the APs will serve at most U_{max} UEs in $\mathcal{D}_l$. The constant U_{max} can also be seen as a type of processing capacity limitation regarding the maximum number of UEs that each AP can process signals. Let $\mathbf{x}_l = \sum_{k=1}^{K} \mathbf{D}_{kl}\mathbf{w}_{kl}s_k$ denote the data signal sent by the AP l. The DL received signal at UE k is given by [27]

$$y_k^{\text{dl}} = \sum_{l=1}^{L} \mathbf{h}_{kl}^{\text{H}}\mathbf{x}_l + n_k = \sum_{l=1}^{L} \mathbf{h}_{kl}^{\text{H}} \left(\sum_{i=1}^{K} \mathbf{D}_{il}\mathbf{w}_{il}s_i \right) + n_k \tag{2.15}$$

where $s_i \in \mathbb{C}$ denotes the signal transmitted for the UE i, which satisfies $\mathbb{E}\left\{\|s_i\|^2\right\} = 1$; $n_k \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma_{\text{dl}}^2\right)$ represents the receiver noise, and $\mathbf{w}_{il} \in \mathbb{C}^{N\times 1}$ is the precoding vector. In a more compact form, (2.15) can be written as

$$y_k^{\text{dl}} = \underbrace{\mathbf{h}_k^{\text{H}}\mathbf{D}_k\mathbf{w}_k s_k}_{\text{Desired signal}} + \sum_{i=1, i\neq k}^{K} \underbrace{\mathbf{h}_k^{\text{H}}\mathbf{D}_i\mathbf{w}_i s_i}_{\text{Interfering signals}} + \underbrace{n_k}_{\text{Noise}}, \tag{2.16}$$

with $\mathbf{w}_i = \left[\mathbf{w}_{i1}^{\text{T}}, ..., \mathbf{w}_{iL}^{\text{T}}\right]^{\text{T}} \in \mathbb{C}^{M\times 1}$ being the collective precoding vector.

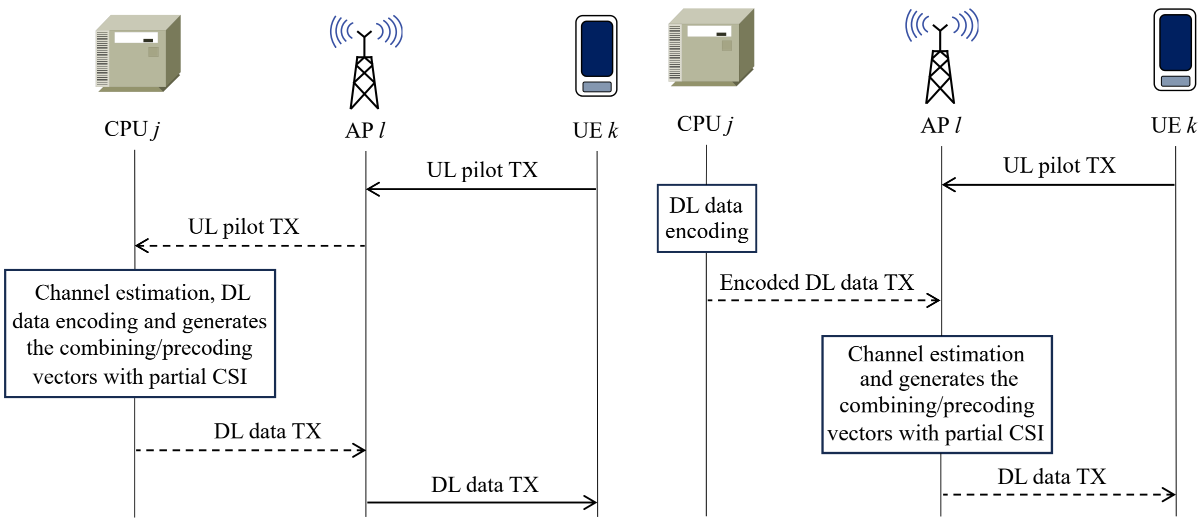
The expression $\mathbf{h}_k^{\text{H}}\mathbf{D}_k\mathbf{w}_k$ in (2.16) is called the effective DL channel and the UE needs to know it to decode the DL data. When only UL pilots are considered, the UE employs the average value of the effective channel to decode the DL data, which can be inaccurate in a distributed network since the channel vector comprises elements that can have quite different channel gains. Thus, other approaches can also estimate the effective channel utilizing DL pilots [51, 52] or blind estimation [45–47]. However, they can reduce the number of samples per data transmission in each coherence block or may rely on coherence blocks with a large number of samples. On the other hand, they can also enhance the network's SE or reduce channel estimation error.

The best strategy will depend on the type of system to be implemented. We utilize UL pilots and rely on reciprocity for UL and DL channels since this approach is widely employed in the literature. Nevertheless, further investigations related to DL pilots or blind estimation may provide valuable insights into the UC literature.

2.2 Network Implementations and Precoding Vectors

UC CF massive MIMO systems are commonly implemented in centralized or distributed manners according to processing capabilities. The centralized implementation places most processing tasks on the CPUs, as Fig. 5 shows. Therefore, the CPUs are responsible for channel estimation, generating the combining and precoding vectors, and processing the DL signals. In the distributed implementation, essential processing functions, such as channel estimation, are moved to the APs. Consequently, the CPUs are only responsible for encoding the DL signals [26]. The centralized implementation usually offers superior interference mitigation since the CPUs can access global CSI, which includes channel estimates and statistics. Conversely, the distributed one can be less complex and avoids the need for transmitting the pilot signals on fronthaul links [28].

Table 1 summarizes the fronthaul signaling required by each network implementation in terms of the number of complex scalars that an AP l sends to a CPU j in each coherence block [26]. One can note that the data traffic (complex scalars for transmitting the UL and DL signals) scales with the number of UEs each AP serves (i.e., $K_l = |\mathcal{D}_l|$) in the distributed implementation. Conversely, in the centralized implementation, the data traffic is proportional to the number of antennas N per AP. In the centralized approach, the APs also transmit UL pilots to the CPUs. This last fact can lead to the idea that distributed implementation always requires less signaling in the fronthaul than the centralized one. However, this is not always true, since if $\tau_c/(\tau_c - \tau_p) \approx 1$ and $K_l > N$, the distributed implementation may require more signaling [28]. It is worth mentioning that this condition is often met in UC systems, since the APs are equipped with few antennas (generally $N \leq 4$). For example, consider a network in which each AP is equipped with $N = 2$ antennas and serves ten UEs simultaneously, i.e., $K_l = 10$. Besides, consider that each coherence block comprises $\tau_c = 200$ complex-valued samples, where $\tau_p = 10$ samples are reserved for pilot signals, $\tau_u = 95$ samples for UL data and $\tau_d = 95$ samples for DL data. In this case, each AP would have to exchange 400 complex scalars with a CPU via fronthaul in each coherence block in the centralized implementation, since $2*(10+95+95) = 400$. In contrast, each AP would need to exchange 1900 complex scalars with a CPU via fronthaul in each coherence block in the distributed implementation, as $10*(95+95) = 1900$.



(a) Centralized (b) Distributed

Figure 5 – Illustration of the two considered network implementations: (a) centralized and (b) distributed. Partial CSI means that the CPUs compute the channel estimates only for the subset of APs serving the UE.

Different combining and precoding techniques can be generated for each network implementation. Notably, MMSE combining represents a signal processing method that minimizes the mean square error (MSE) of data detection and maximizes the SINR for all UEs in the network. It relies on channel estimates available on the CPUs, effectively

Table 1 – Number of complex scalars that an AP l has to exchange with a CPU j via fronthaul in each coherence block in the centralized and distributed implementations.

	Pilot signals	UL signals	DL signals
Centralized	$\tau_p N$	$\tau_u N$	$\tau_d N$
Distributed	-	$\tau_u K_l$	$\tau_d K_l$

mitigating interference. The MMSE combining can be written as

$$\mathbf{v}_k^{\mathrm{MMSE}} = \eta_k \left(\sum_{i=1}^{K} \eta_i \mathbf{D}_k \left(\widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} + \mathbf{C}_i \right) \mathbf{D}_k + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k, \tag{2.17}$$

where $\mathbf{C}_k$ is the correlation matrix of the estimation error $\mathbf{h}_{ek} = \mathbf{h}_k - \widehat{\mathbf{h}}_k$, such that $\mathbf{C}_k = \mathbb{E}\{\mathbf{h}_{ek}\mathbf{h}_{ek}^{\mathrm{H}}\}$. For the centralized implementation, the combining vector is generated using the collective channel vector, given by $\mathbf{h}_k = \left[\mathbf{h}_{k1}^{\mathrm{T}}, ..., \mathbf{h}_{kL}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$. Additionally, it utilizes the diagonal block matrix $\mathbf{D}_k = \mathrm{diag}\left(\mathbf{D}_{k1}, ..., \mathbf{D}_{kL}\right) \in \mathbb{C}^{M\times M}$.

The distributed version of MMSE combining is named local MMSE (L-MMSE), which utilizes the local channel estimates of all UEs in the network, which is given by

$$\mathbf{v}_{kl}^{\mathrm{L-MMSE}} = \eta_k \left(\sum_{i=1}^{K} \eta_i \left(\widehat{\mathbf{h}}_{il} \widehat{\mathbf{h}}_{il}^{\mathrm{H}} + \mathbf{C}_{il} \right) + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{N} \right)^{-1} \mathbf{D}_{kl} \widehat{\mathbf{h}}_{kl}. \tag{2.18}$$

Both MMSE and L-MMSE are not scalable combining techniques since they need the channel estimations of all UEs in the network. Therefore, some adjustments must be made to provide scalability to these schemes.

The partial MMSE (P-MMSE) is the scalable version of MMSE combining. The only difference between them is that the P-MMSE scheme does not consider all UEs' channel estimates but only the subset of UEs that most interfere with UE k. Specifically, the subset of UEs partially served by the same APs as UE k, which is computed as

$$\mathcal{P}_k = \left\{ i : \mathbf{D}_k \mathbf{D}_i \neq \mathbf{0}_{LN\times LN} \right\}. \tag{2.19}$$

Hence, the P-MMSE combining can be expressed as [27]

$$\mathbf{v}_k^{\mathrm{P-MMSE}} = \eta_k \left(\sum_{i\in\mathcal{P}_k} \eta_i \mathbf{D}_k \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} \mathbf{D}_k + \mathbf{Z}_{\mathcal{P}_k} + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k \tag{2.20}$$

where

$$\mathbf{Z}_{\mathcal{P}_k} = \sum_{i\in\mathcal{P}_k} \eta_i \mathbf{D}_k \mathbf{C}_i \mathbf{D}_k. \tag{2.21}$$

The local partial minimum mean-squared-error (LP-MMSE) is the scalable version of L-MMSE combining scheme. It considers only the channel estimates of the UEs that AP l serves, i.e., $k \in \mathcal{D}_l$, since $K_l = |\mathcal{D}_l|$ is limited. The LP-MMSE combining can be written as

$$\mathbf{v}_{kl}^{\mathrm{LP-MMSE}} = p_k \left(\sum_{i\in\mathcal{D}_l} p_i \left(\widehat{\mathbf{h}}_{il} \widehat{\mathbf{h}}_{il}^{\mathrm{H}} + \mathbf{C}_{il} \right) + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{N} \right)^{-1} \mathbf{D}_{kl} \widehat{\mathbf{h}}_{kl}. \tag{2.22}$$

There are also two other techniques that are commonly used in UC CF massive MIMO literature for combining signals, which are called maximum-ratio (MR) and partial regularized zero-forcing (P-RZF). The MR (or conjugated beamforming) is the most common choice for the distributed implementation, which is expressed as

$$\mathbf{v}_{kl}^{\mathrm{MR}} = \widehat{\mathbf{h}}_{kl}. \tag{2.23}$$

The MR combining is a scalable and low-complexity technique. It maximizes the ratio between the received power and the square norm of the UL combining vector but cannot efficiently mitigate interference among UEs.

The P-RZF is utilized in centralized implementation and simplifies the P-MMSE combining aiming to decrease the CC. It neglects the matrix $\mathbf{Z}_{\mathcal{P}_k}$ by observing that when the channel conditions of the interfering UEs in $\mathcal{P}_k$ are good, all the corresponding estimation error correlation matrices $\mathbf{C}_i$ will be small. This modification has a negligible influence on CC but allows to rewrite the P-RZF combining vector as [27]

$$\mathbf{v}_{k}^{\mathrm{P-RZF}} = \left[\mathbf{D}_k \widehat{\mathbf{H}}_{\mathcal{P}_k} \left(\widehat{\mathbf{H}}_{\mathcal{P}_k}^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{H}}_{\mathcal{P}_k} + \sigma_{\mathrm{ul}}^2 \mathbf{P}_{\mathcal{P}_k}^{-1}\right)^{-1}\right]_{:,1} \tag{2.24}$$

where $[\cdot]_{:,1}$ represents the operation of only keeping the first column of its matrix argument, $\widehat{\mathbf{H}}_{\mathcal{P}_k} \in \mathbb{C}^{LN \times |\mathcal{P}_k|}$ contains the stacked vectors $\widehat{\mathbf{h}}_i$ with indices $i \in \mathcal{P}_k$, with the first column being $\widehat{\mathbf{h}}_k$, and $\mathbf{P}_{\mathcal{P}_k} \in \mathbb{R}^{|\mathcal{P}_k| \times |\mathcal{P}_k|}$ is a diagonal matrix containing the transmit powers η_i for $i \in \mathcal{P}_k$, listed in the same order as the columns $\widehat{\mathbf{H}}_{\mathcal{P}_k}$. Basically, the P-RZF combining takes the pseudo-inverse of the partial channel $\widehat{\mathbf{H}}_{\mathcal{P}_k}$ and regularizes it by adding the matrix $\sigma_{\mathrm{ul}}^2 \mathbf{P}_{\mathcal{P}_k}^{-1}$. Combining vectors using pseudo-inverses forces the interference between the UEs to zero. However, it may result in notable reductions in the desired signal power when the UEs possess similar channels.

Motivated by the UL-DL duality, the DL precoding vectors can be select based on the UL combiners [26, 27]. This strategy also reduces CC for computing the precoding vectors, since it is only needed to normalize the combining vectors, such as

$$\mathbf{w}_k = \sqrt{\varrho_k} \frac{\mathbf{v}_k}{\sqrt{\mathbb{E}\left\{\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{v}_k\right\}}}, \quad \mathbf{w}_{kl} = \sqrt{\varrho_{kl}} \frac{\mathbf{v}_{kl}}{\sqrt{\mathbb{E}\left\{\mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{v}_{kl}\right\}}}, \tag{2.25}$$

where $\mathbf{v}_k$ and $\mathbf{v}_{kl}$ represent the arbitrary combining vector used for centralized and distributed implementations, respectively. Furthermore, the terms ϱ_k and ϱ_{kl} denote the transmission powers assigned to the UE k in centralized and distributed implementations. The term ϱ_k represents the total transmission power assigned to UE k from all its serving APs, while ϱ_{kl} stands for the transmission power that the UE k receives from AP l.

We adopt the MR and LP-MMSE precoding for the distributed implementation, while we employ the P-RZF and P-MMSE for the centralized one. Besides, note that all precoding vectors are calculated under imperfect CSI. For instance, $\mathbf{v}_k = \widehat{\mathbf{h}}_k$ and $\mathbf{v}_{kl} = \widehat{\mathbf{h}}_{kl}$

for the MR precoding. Table 2 exhibits the CC (in number of complex multiplications) required to compute the precoding schemes for each UE in each coherence block. Note that although the CC does not grow with the number of UEs in a scalable system (i.e., K_l is restricted by U_{max}), it can still increase with the number of APs per UE (L_k).

Table 2 – Number of complex multiplications required to generate the combining vectors for each UE in each coherence block [26].

Scheme	Channel estimation	Combining vector computation
MMSE	$(N\tau_p + N^2)\, KL_k$	$\frac{(NL_k)^2+NL_k}{2}K + (NL_k)^2 + \frac{(NL_k)^3-NL_k}{3}$
P-MMSE	$(N\tau_p + N^2)\, \lvert\mathcal{P}_k\rvert\, L_k$	$\frac{(NL_k)^2+NL_k}{2}\lvert\mathcal{P}_k\rvert + (NL_k)^2 + \frac{(NL_k)^3-NL_k}{3}$
LP-MMSE	$(N\tau_p + N^2)\sum_{l\in\mathcal{M}_k} K_l$	$\frac{N^2+N}{2}\sum_{l\in\mathcal{M}_k} K_l + \left(\frac{N^3-N}{3} + N^2\right) L_k$
MR	$(N\tau_p + N^2)\, L_k$	-

2.3 Power Allocation

Two heuristic methods for power allocation are considered to address scalability aspects in both network implementations. We employ fractional power allocation[2] for the centralized one, since it performs better than equal power allocation for the worst UEs. In the centralized implementation, the transmission powers that the APs assign to the UE k are coupled by means of the precoding vector in (2.25). Thus, it is usual to compute the total transmission power that each UE receives from all its serving APs first, i.e., ϱ_k. Then, the normalized precoding vector in (2.25) determines how this power is distributed between the APs serving the UE k (i.e., for all l in $\mathcal{M}_k$).

This is done to maintain the same direction of the precoding vector. Otherwise, the capacity of the APs to mitigate each others' interference can be lost in centralized implementation. The DL fractional power allocation is computed as [27]

$$\varrho_k = \varrho_d \frac{\left(\sum\limits_{l\in\mathcal{M}_k} \beta_{kl}\right)^{v'} \omega_k^{-\kappa'}}{\max\limits_{\ell\in\mathcal{M}_k} \sum_{i\in\mathcal{D}_\ell} \left(\sum\limits_{l\in\mathcal{M}_i} \beta_{il}\right)^{v'} \omega_i^{1-\kappa'}}, \tag{2.26}$$

with ω_k being calculated as

$$\omega_k = \max_{\ell\in\mathcal{M}_k} \mathbb{E}\left\{\lVert\bar{\mathbf{w}}_{k\ell}\rVert^2\right\} \tag{2.27}$$

where $\bar{\mathbf{w}}_{k\ell} \triangleq \mathbf{v}_{k\ell} \in \mathbb{C}^{N\times 1}$ is the fraction of the centralized collective precoding vector ($\bar{\mathbf{w}}_k \triangleq \mathbf{v}_k \in \mathbb{C}^{LN\times 1}$) that correspond to AP ℓ and the term ϱ_d represents the total transmission power of each AP. The normalization factor in the denominator of (2.26) is utilized to ensure that none of the APs will exceed the maximum transmission power

[2] There is a wide range of power allocation methods in the literature [37–43]. However, analyses involving the best power allocation algorithm are out of the scope of this thesis.

ϱ_d [27]. Thus, ϱ_k is a function of β_{kl}, $\upsilon' \in [-1,1]$, $\kappa' \in [0,1]$, and β_{kl}, which is large-scale fading of UE k regarding AP l. Moreover, υ' and κ' are project parameters [27].

For the distributed implementation, we employ a method that also divides the power resources proportionally to the large-scale fading gains of each UE [61]. Thus, ϱ_{kl} is given by

$$\varrho_{kl} = \begin{cases} \varrho_l \frac{\sqrt{\beta_{kl}}}{\sum_{i\in\mathcal{D}_l}\sqrt{\beta_{il}}} & \text{if } k \in \mathcal{D}_l \\ 0 & \text{otherwise} \end{cases}, \tag{2.28}$$

where $\varrho_l = \varrho_d$ and $\beta_{kl} = \text{tr}\left(\mathbf{R}_{kl}\right)/N$ if the channel statistics are assumed to be perfectly known or, otherwise, $\beta_{kl} = \text{tr}(\widehat{\mathbf{R}}_{kl})/N$.

2.4 Spectral and Energy Efficiencies

In order to calculate the SE of DL channels, we rely on the received signal presented in (2.16). Thus, an achievable DL SE for the UE k can be expressed as [27]

$$\text{SE}_k = P_f \log_2\left(1 + \text{SINR}_k\right), \tag{2.29}$$

where P_f is the pre-log factor, which is a fraction of samples per coherence block that is used to transmit data. For perfect knowledge of correlation matrices, $P_f = \tau_d/\tau_c$ and $P_f = 1 - (\tau_p/\tau_c) - \alpha$ for imperfect knowledge, where $\alpha = N_R\tau_p K/\tau_s\tau_c$. The term SINR_k denotes the DL SINR. From (2.16), the SINR_k can be computed as

$$\text{SINR}_k = \frac{\left|\text{DS}_k\right|^2}{\text{IS}_k - \left|\text{DS}_k\right|^2 + \sigma^2_{\text{dl}}}. \tag{2.30}$$

where $\text{DS}_k = \mathbb{E}\left\{\mathbf{h}_k^{\text{H}}\mathbf{D}_k\mathbf{w}_k\right\}$ denotes the desired signal and $\text{IS}_k = \sum_{i=1}^{K}\mathbb{E}\left\{\left|\mathbf{h}_k^{\text{H}}\mathbf{D}_i\mathbf{w}_i\right|^2\right\}$ stands for the interference. Eq. (2.29) is also known as hardening bound, which is commonly used in massive MIMO theory and is valid for any choice of precoding vectors [28, 37]. It can be seen as a capacity lower bound and, unfortunately, it does not have a closed-form expression when using P-MMSE and LP-MMSE schemes, but can be computed through Monte-Carlo simulations if $\mathbf{w}_i$ is selected as in (2.25). Besides, all expectations presented in (2.30) are related to the channel realizations [27].

The total EE in bit/Joule is defined as the ratio between the total data rate $R_t = B_s\sum_{k=1}^{K}\text{SE}_k$, and the total power consumption in Watts [97]. In UC systems, the EE can be interpreted as the total data rate that the network can achieve for a given power consumption. We have computed the EE as [24]

$$\text{EE}_t = \frac{R_t}{\sum\limits_{l=1}^{L}\left\{\frac{1}{\nu_l}\mathbb{E}\left\{\|\mathbf{x}_l\|^2\right\} + NP_{\text{tc},l} + P_{\text{fh},l}\right\}}, \tag{2.31}$$

where $0 < \nu_l \leq 1$ denotes the efficiency of the power amplifier in the AP l, and $P_{\text{tc},l}$ is the power that each antenna of AP l needs to run internal components, such as converters and filters. Besides, $P_{\text{fh},l}$ is the power consumption in the fronthaul link connecting a CPU to AP l, which is calculated as

$$P_{\text{fh},l} = P_{0,l} + P_{\text{ft},l}\, B_s \sum_{k \in \mathcal{D}_l} \text{SE}_k, \tag{2.32}$$

where $P_{0,l}$ is a fixed power consumption of each link and $P_{\text{ft},l}$ is the traffic-dependent power in Watt per bit/s. It is noteworthy that (2.32) is only suitable for distributed network implementations since the term $\sum_{k \in \mathcal{D}_l} \text{SE}_k$ scales with the subset of UEs that the AP l serves ($\mathcal{D}_l$). In a centralized implementation, (2.32) should be proportional to the number of antennas N in each AP. Besides, several other factors may affect the EE of UC systems, such as the power consumption due to CC in APs and CPUs, idle mode, and others. Such an investigation comprises the next steps of the research involving this thesis.

3 Matched-Decision AP Selection

The literature presents several approaches for AP selection in UC CF massive MIMO systems, which associate UEs and APs to improve some network performance metrics, such as EE, SE, and others. However, they do not evaluate whether these connections benefit both the UEs and APs. They generally consider that the AP selects a subset of UEs to serve or that the UE selects a subset of APs to connect but do not consider a matched-decision between the most suitable connections for both the UEs and APs. These strategies may also not prevent the UEs from being dropped from the network and may fail in scalability since they do not limit the number of UEs each AP can serve. Additionally, these AP selection methods may associate the UEs with APs that contribute only marginally to UE's performance, leading to ineffective use of network resources.

This chapter presents the general AP selection framework proposed in this thesis that allows a matched-decision between the most advantageous connections for UEs and APs. The proposed framework is flexible and can be adapted to different AP selection criteria. Furthermore, it can provide scalability for unscalable AP selection schemes while ensuring the connections of all UEs to the network. This chapter also presents the three fine-tuning schemes proposed to drop UE-AP connections that contribute only marginally to the system performance.

3.1 Proposed AP Selection Framework

This section presents a novel AP selection framework that exploits a competitive mechanism and considers a matched-decision among UEs and APs. The scheme is divided into two stages, and the flowchart exhibited in Fig. 6 provides an overview of the method's operation. The matched-decision process occurs in the first stage (named intermediate AP cluster). In this stage, the UEs connect to an intermediate subset of APs, aiming to make the UEs and APs establish the best connection for both in terms of large-scale fading. In the second stage (called final AP cluster), the UEs try to connect to more APs and expand their AP clusters, intending to improve the SE. For clarification, Figs. 7 (a) and 7 (b) illustrate the intermediate and final AP clusters, respectively. The first stage can enable better use of the power resources since the matched-decision allows the APs to serve the best UEs in their vicinity. The second one gives the worst UEs a chance to increase their SE.

We assume a specific limitation in the processing capacity of the APs, that is, each AP can serve only a limited number of UEs, named U_{max}. However, unlike previous works, we do not consider that U_{max} is always equal to τ_p [26, 27]. Instead, we assume that U_{max}

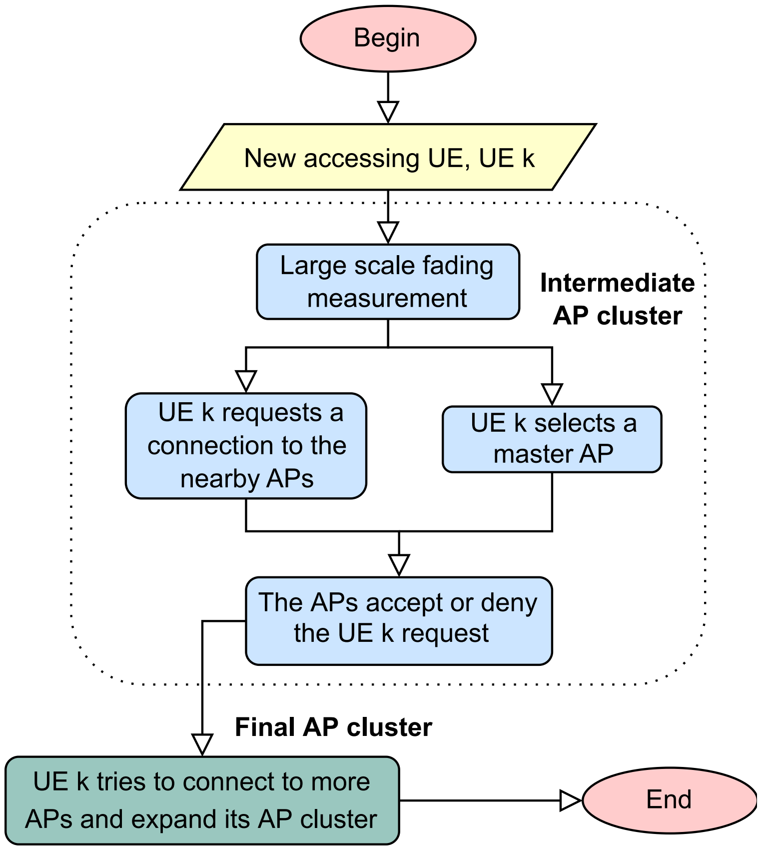
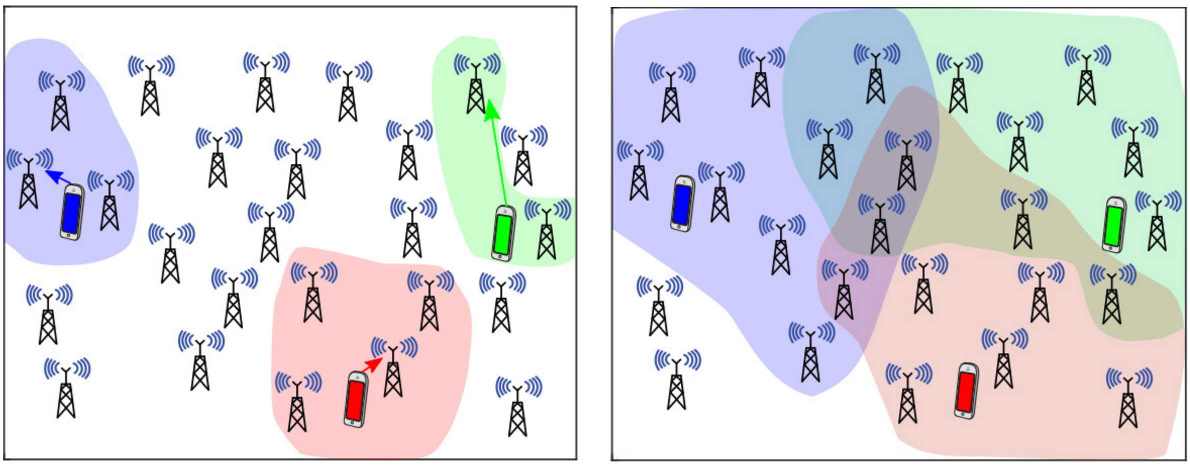


Figure 6 – Flowchart of the proposed matched-decision AP selection method.



(a) Intermediate stage (b) Final stage

Figure 7 – Illustration of the AP cluster formation. (a) Intermediate AP clusters, with arrows indicating the master APs. (b) Final AP clusters after the second stage.

depends on the AP processing capabilities, such that $1 \leq U_{max} \leq \tau_p$. Additionally, the AP selection is generated on a per-UE basis to achieve scalability, such that the AP clustering occurs only among the UEs and the APs. The CPUs do not participate in the AP cluster formation.

3.1.1 Intermediate AP Cluster

Fig. 7 (a) illustrates the intermediate AP clusters created for each UE in the network. Its generation is detailed as follows: when a new UE k enters the network, it measures the large-scale fading β_{kl} of the APs in its vicinity. Then, the UE requests a connection to a subset of APs according to a decision criterion that follows the requirements of the system's design. This thesis considers that the UE k will request a connection only to the APs whose channel gains satisfy $\beta_{kl} > \gamma$. The target is to evaluate the system performance under a controlled parameter (γ), where γ refers to a threshold gain. The subset of APs selected by the UE k is denoted by $\mathbf{e}_k = [\mathrm{e}_{k1}, ..., \mathrm{e}_{kL}] \in \mathbb{N}^{1\times L}$ and is defined as

$$\mathrm{e}_{kl} = \begin{cases} 1 & \text{if } \ \beta_{kl} > \gamma \\ 0 & \text{otherwise} \end{cases}. \tag{3.1}$$

However, although UE k desires to connect to the APs that meet the decision criterion, the connection will only be carried out whether these APs individually accept the request of UE k. That is, there must be a matched-decision between the UE k and the APs. In other words, the connection must be advantageous for both UE k and APs. The APs can employ several decision criteria, such as the least pilot contamination [26], effective channel gain [62], among others. Nonetheless, to use similar criteria across UEs and APs, we assume that the decisions rely on channel gain β_{kl} in the APs. The decisions are denoted by $\mathbf{f}_k = [\mathrm{f}_{k1}, ..., \mathrm{f}_{kL}] \in \mathbb{N}^{1\times L}$ and can be summarized as

$$\mathrm{f}_{kl} = \begin{cases} 1 & \text{if } \ \beta_{kl} > \beta_{il}^{min} \\ 0 & \text{otherwise} \end{cases}, \tag{3.2}$$

where $i \neq k$ denotes the UE with the smallest channel gain (β_{il}^{min}) that the AP l serves in $\mathcal{D}_l$. Recall that $\mathcal{D}_l$ denotes the subset of UEs served by AP l, with $K_l = |\mathcal{D}_l|$. In case of $\beta_{kl} > \beta_{il}^{min}$, the AP accepts the UE k and keeps the connection of the UE i only if there are available connections (i.e., $K_l < U_{max}$). The intermediate AP cluster is given by

$$\mathbf{c}_k^{int} = \mathbf{e}_k \wedge \mathbf{f}_k, \tag{3.3}$$

where $\wedge$ is the logical operation AND. From (3.3), one can note that the UE k may not connect with any AP if it is rejected by all APs in (3.2). To circumvent this issue, the UE k also claims a master AP. This AP ensures that the network will no longer drop the UE k since it serves the UE regardless of its channel condition [26]. We consider that each UE has a master AP and that its choice is independent of the threshold γ. Hence, the UE connects to at least one AP. Consequently, the cardinality of the subset of APs that serve the UE k (i.e., $\mathcal{M}_k$) becomes

$$L_k = \sum_{l\in\mathcal{M}_k} \mathrm{c}_{kl}^{int} \geq 1. \tag{3.4}$$

In Fig. 7 (a), the arrows indicate the master AP of each UE. The proposed strategy assumes that the procedure for choosing a master AP is the same for all UEs. Thus, the procedure will be explained only for the UE k. To choose a master AP, the UE k solicits a connection to the available APs. Then, the available APs respond, and the UE k selects the one with the strongest channel gain β_{kl} to be its master AP. Let $\mathcal{A}_l \subset \mathcal{D}_l$ denote the subset of UEs that the AP l is serving as a master. The available APs are the ones presenting $|\mathcal{A}_l| < U_{max}$, $\forall l \in \{1, \ldots, L\}$. That is, these APs are not using their total processing capacity only for serving UEs as master APs. Therefore, the master AP of UE k can be defined by

$$\begin{array}{l} l = \arg\max_l \beta_{kl} \\ \text{s.t} \, |\mathcal{A}_l| < U_{max} \end{array}, \tag{3.5}$$

which means that although several APs (that present $|\mathcal{A}_l| < U_{max}$) reply to the request of the UE k, it selects only the AP with the highest β_{kl} to be its master AP. It is worth noting that in (3.5), $|\mathcal{A}_l| + |\mathcal{B}_l| \leq U_{max}$, where $\mathcal{B}_l \subset \mathcal{D}_l$ denotes the subset of UEs that the AP l serves, but not as a master (i.e., UEs that the AP can drop). Due to the addition of the master AP, (3.1) becomes

$$\mathrm{e}_{kl} = \begin{cases} 1 & \text{if } (\text{UE k} \in \mathcal{A}_l) \vee (\beta_{kl} > \gamma) \\ 0 & \text{otherwise} \end{cases}, \tag{3.6}$$

where $\mathcal{A}_l \subset \mathcal{D}_l$ and $\vee$ represents the logical operation OR. Additionally, (3.2) changes to

$$\mathrm{f}_{kl} = \begin{cases} 1 & \text{if } (\text{UE k} \in \mathcal{A}_l) \vee (\beta_{kl} > \beta_{il}^{\min}) \\ 0 & \text{otherwise} \end{cases}, \tag{3.7}$$

where $i \neq k$ and $i \in \mathcal{B}_l$. One can note that the network will no longer drop the UE k if we apply (3.6) and (3.7) in (3.3) instead of (3.1) and (3.2), since the choice of the master AP is independent of γ. It is worth noting that the subset $\mathcal{B}_l$ does not affect the master AP assignment in (3.5). For instance, if $K_l = U_{max}$ and $|\mathcal{B}_l| \geq 1$, the AP l could drop the UE with the weakest channel gain belonging to subset $\mathcal{B}_l$ (UEs that the AP l serves, but not as a master) in order to be the master of UE k in subset $\mathcal{A}_l$. Such an approach guarantees that $K_l \leq U_{max}$ since $|\mathcal{A}_l| + |\mathcal{B}_l| \leq U_{max}$. Algorithm 1 summarizes the intermediate AP selection process.

3.1.2 Final AP Cluster Formation

Fig. 7 (b) depicts the final AP clusters of each UE. Its generation is detailed as follows: after the formation of vectors $\mathbf{e}_k$ and $\mathbf{f}_k$, the UE k may have been rejected by some APs. However, the UE has a new chance to link to these APs and then expand its AP cluster in this second step. To do this, an AP l' that initially rejected the UE k, verify if there are still connections available for a new UE. That is, it checks if $K_{l'} < U_{max}$. Then,

Algorithm 1: Intermediate AP cluster

Input: γ, U_{max};

The UE k connects to a master AP by solving (3.5);

Update $\mathcal{A}_l$ according to the solution of (3.5);

for $l = 1$ **to** L **do**

  $\mathrm{e}_{kl} = 0$; $\mathrm{f}_{kl} = 0$; $\mathrm{c}_{kl} = 0$;

  // The UE k requests connections to the nearby APs:

  if $k \in \mathcal{A}_l$ **or** $\beta_{kl} > \gamma$ **then**

    $\mathrm{e}_{kl} = 1$;

  end

  // The APs accept or reject the UE request:

  if $k \in \mathcal{A}_l$ **or** $\beta_{kl} > \beta_{il}^{min}$ **then**

    $\mathrm{f}_{kl} = 1$; $\mathrm{f}_{il} = 1$; // where $i \in \mathcal{B}_l$

    if $K_l = \mathrm{U}_{max}$ **then**

      $\mathrm{f}_{il} = 0$;

    end

  end

  $\mathrm{c}_{kl}^{int} = (\mathrm{e}_{kl} \wedge \mathrm{f}_{kl})$ // Matched-decision

end

Output: $\mathbf{c}_k^{int} = [\mathrm{c}_{k1}^{int}, \ldots, \mathrm{c}_{kL}^{int}]$.

the AP accepts the request of the UE k if this condition is satisfied. Computationally, this can be represented by the vector $\mathbf{z}_{kl'} = [\mathrm{z}_{k1}, ..., \mathrm{z}_{kL}] \in \mathbb{N}^{1\times L}$, which is given by

$$\mathrm{z}_{kl'} = \begin{cases} 1 & \text{if } K_{l'} < U_{max} \\ 0 & \text{otherwise} \end{cases}, \tag{3.8}$$

where $l' \neq l$. However, the UE k may be dropped again in (3.7) if a new UE with a better channel condition enters the network. In the end, the final AP cluster is computed as

$$\mathbf{c}_k = \mathbf{c}_k^{int} \vee \mathbf{z}_k, \tag{3.9}$$

where the operator $\vee$ denotes the logical operation OR. From (3.9), one can compute the number of APs serving the UE k by calculating $L_k = \sum_{l\in\mathcal{M}_k} \mathbf{c}_k$. One can also compute $\mathbf{D}_k$ through $\mathbf{c}_k$ by assuming that $\mathbf{D}_{kl} = \mathbf{I}_N$, when $\mathrm{c}_{kl} = 1$. Otherwise $\mathbf{D}_{kl} = \mathbf{0}_N$, for $l = \{1, ..., L\}$. For clarification, Algorithm 2 summarizes the final AP clustering.

The AP selection proposed in this section is composed of the Algorithms 1 and 2. In the first step of the Algorithm 1, the time complexity for each new UE to select its master AP by solving (3.5) is $\mathcal{O}(L)$. Then, the complexity for each new UE that requests connections to nearby APs is $\mathcal{O}(|\mathcal{A}_l|)$, while for each AP to accept or reject the request is $\mathcal{O}(|\mathcal{B}_l|)$. This comes from the fact that only the UEs belonging to subsets $\mathcal{A}_l$ and $\mathcal{B}_l$ are evaluated. The complexity for computing the intermediate AP cluster is $\mathcal{O}(L(|\mathcal{A}_l| + |\mathcal{B}_l|))$, since these steps are repeated for each AP. Then, noticing that $|\mathcal{A}_l| + |\mathcal{B}_l| \leq K_l$, the complexity of Algorithm 1 simplifies to $\mathcal{O}(LK_l)$. The complexity to compute the final

Algorithm 2: Final AP cluster

Input: $\mathbf{c}_k^{int}$, U_{max};

for $l = 1$ **to** L **do**

  // Connect the k-th UE with available APs:

  if $|\mathcal{D}_l| < U_{max}$ **then**

    $\mathrm{z}_{kl} = 1$;

  end

  $\mathrm{c}_{kl} = \mathrm{c}_{kl}^{int} \vee \mathrm{z}_{kl}$;

end

Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

AP cluster in Algorithm 2 is also $\mathcal{O}(LK_l)$, since only the UEs belonging to subset $\mathcal{D}_l$ are evaluated for each AP. Therefore, the overall time complexity of the proposed AP selection is $\mathcal{O}(LK_l)$, which can be further simplified to $\mathcal{O}(LU_{max})$ by noticing that $K_l \leq U_{max}$. Therefore, one can note the scalability aspect of the proposed strategy since its complexity does not increase with K.

It is noteworthy note that the decision criteria employed in this section rely on large-scale fading coefficients. This is because these statistics remain valid for several coherence blocks, which means that we do not need to re-run the algorithm too often. However, the decision criteria could also include other metrics, such as pilot contamination and EE. The purpose of the matched-decision algorithm is to generate a compromise between the best connections for UEs and APs. Therefore, it can be generalized to other metrics. Furthermore, although this is not presented in this thesis, the concept of matched-decision can also be extended to CPUs. In this scenario, one should analyze the most advantageous connections for UEs, APs, and CPUs.

3.2 Comparison with other AP Selection Methods

The proposed solution aims to generate AP clusters composed of the more convenient connections for UEs and APs. Consequently, it inherits several characteristics of classical AP selection schemes (decisions taken only in the UEs or in the APs) and can degenerate into these by adjusting the vectors $\mathbf{e}_k$, $\mathbf{f}_k$, and $\mathbf{z}_k$. Before showing it, let us briefly describe some solutions that we use for comparisons.

3.2.1 Brief Description of Baseline AP Selection Methods

- **Canonical CF:** it is a NS scheme in which the AP cluster of each UE is composed of all APs. This method improves the SE of the worst UEs and increases the network's coverage probability compared to cellular networks [23].

- **UCC:** the AP serves the U_{max} UEs with the greatest estimated channel in each coherence block. We have adjusted this strategy to consider only the large-scale fading to avoid performing AP selection in each coherence block. Thus, it is considered that the AP serves the U_{max} UEs presenting the largest large-scale fading in their vicinity. The user-centric clustering (UCC) is a NS method that does not prevent the worst UEs from being dropped. The time complexity of this method is $\mathcal{O}(LK \log K)$ for each UE as whenever a new UE enters the network, all APs have to perform a sorting operation to select U_{max} UEs to serve [25].

- **LSFB:** the UE establishes a connection with the subset of APs that contribute most to the sum of its total channel gain, in percentage $\delta\%$. This strategy has complexity $\mathcal{O}(L \log L)$ for each UE, as each new UE sorts the channel gains of L APs during the AP selection process. One can note that this method could be scalable, as its time complexity does not increase with K. Nevertheless, the largest-large-scale-fading-based (LSFB) is also a NS scheme since it does not limit the number of UEs that each AP can serve [24].

- **Scalable CF:** the UE enters the network and connects to a master AP. Then, the master AP assigns a pilot t_k to the UE and informs other APs in its neighborhood (named non-master APs) that it is serving the new UE in the pilot t_k. Posteriorly, Each non-master APs individually decides to serve or not the UE. To be served, the UE has to present a channel gain above a threshold value. Then, one of the following conditions has to be met: i) the AP is available for another UE, ii) The UE's channel gain must be greater than the channel gain of the UE that is already using the pilot t_k at the non-master AP. Each AP serves only one UE per pilot. The first τ_p UEs are assigned to mutually orthogonal pilots, and the remaining ones to the pilot causing the lowest pilot contamination in the master AP. Among the reference schemes analyzed, this is the only one that is scalable and prevents the worst UEs from being dropped. However, providing a mechanism that aims further to improve the SE of the worst UEs is not one of the goals of this method. The time complexity is $\mathcal{O}(L\tau_p)$ for each UE, which is due to the fact that after the pilot assignment, each AP chooses to serve up to one UE per pilot. The time complexity of this pilot assignment method for all UEs K is $\mathcal{O}((K - \tau_p)\tau_p)$ [26].

3.2.2 Relationships Between the Matched-Decision and Baseline Methods

The matched-decision scheme behaves like a UCC strategy if we set $\mathrm{e}_{kl} = 1$ for all APs, which is similar to considering a small γ in (3.6). By doing that, the UEs' choices do not impact (3.3), and the APs' decisions dominate the AP cluster formation. Consequently, the APs tend to select the UEs presenting the best channel gain in their vicinity in (3.7), leading to a UCC scheme. However, this UCC implementation achieves scalability, guarantees connection for all UEs, and its time complexity is independent of the number of UEs. Hereafter, we name it matched-decision (MD) UCC.

The proposed scheme can also behave like a scalable version of the LSFB algorithm. Hereafter, we name it as MD LSFB. To this end, one should consider $\mathrm{e}_{kl} = 1$ for the subset of APs that contribute most to the total sum channel gain $\delta\%$, instead of relying on γ in (3.6). One can also achieve a similar implementation of the scalable CF scheme by considering $\mathrm{e}_{kl} = 1$ and $\mathrm{z}_{kl} = 0$ for all APs. Then, (3.7) should be modified to make the APs serve only the UEs, causing the least pilot contamination to the received signal. Regarding the canonical CF, it is only a particular case of the UC approach when $\mathrm{c}_{kl} = 1$ and $\mathbf{D}_{kl} = \mathbf{I}_N$ for all APs and UEs [26].

3.3 Proposed Fine-Tuning AP Selection Schemes

The AP cluster of each UE created by AP selection schemes can comprise APs that contribute only marginally to the UE's performance. Therefore, this section proposes two strategies to reduce the number of APs connected to each UE while avoiding reducing the SE significantly, and a third one that aims to improve the EE. We have named these solutions as fine-tuning AP selection.

3.3.1 Fine-Tuning Based on Allocated Power

The first one is performed locally in each AP, without the CPUs participation. The AP l can drop not mastered UEs (i.e., UEs that are in $\mathcal{B}_l$) that receive only a small fraction of the total power in (2.28). For this, the AP sorts $\{\varrho_{1l}, ..., \varrho_{kl}\}$ in descending order to identify the not mastered UEs that receive more power, leading to $\{\bar{\varrho}_{1l}, ..., \bar{\varrho}_{k'l}\}$. The indexes of the unsorted UEs are stored in the k'-th element of subset $\bar{\mathcal{B}}_l$. Then, the AP carries out a cumulative sum, which can be expressed as

$$\varrho_{k'l}^{\mathrm{sum}} = \begin{cases} \frac{\bar{\varrho}_{k'l}}{\sum_{k\in\mathcal{D}_l}\varrho_{kl}} & \text{if } k' = 1 \\ \frac{\bar{\varrho}_{k'l}}{\sum_{k\in\mathcal{D}_l}\varrho_{kl}} + \varrho_{(k'-1)l}^{\mathrm{sum}} & \text{otherwise} \end{cases}. \tag{3.10}$$

After that, the AP makes $\mathrm{c}_{kl} = 1$ for the UEs that contribute to at least $\Gamma\%$ of the cumulative sum in (3.10) and $\mathrm{c}_{kl} = 0$ for the remaining ones. Algorithm 3 summarizes the entire process. A similar approach has been considered in [24], where an AP selection is carried out based on each UE's received power after acquiring power coefficients from an optimal power allocation strategy that maximizes EE. However, such an approach is NS as its complexity increases with K. Besides, it does not use the power to fine-tune the AP clusters but only generates them.

3.3.2 Fine-Tuning Based on Spectral Efficiency

The second fine-tuning scheme works in a centralized fashion, and it is based on SE. Specifically, the CPUs drop the connections of APs that contribute only marginally to the SE of the UE k. For fine-tuning the AP clusters, the CPUs consider that all APs apply MR precoding and assume perfect CSI to simplify the calculation of the desired signal (DS_k) and interference (IS_k) terms in (2.30), which are computed as

$$\begin{aligned}\mathrm{DS}_k &= \mathbb{E}\left\{\mathbf{h}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{w}_k\right\} = \sum_{l\in\mathcal{M}_k}\sqrt{\varrho_k\,\mathrm{tr}\left(\mathbf{R}_{kl}\right)}\\ \mathrm{IS}_k &= \sum_{i\in\mathcal{P}_k}\mathbb{E}\left\{\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{D}_i\mathbf{w}_i\right|^2\right\} = \sum_{i\in\mathcal{P}_k}\varrho_i\sum_{l\in\mathcal{M}_i}\frac{\mathrm{tr}\left(\mathbf{R}_{il}\mathbf{R}_{kl}\right)}{\mathrm{tr}\left(\mathbf{R}_{il}\right)},\end{aligned} \tag{3.11}$$

where $\mathcal{P}_k$ denotes the subset of UEs that are partially served by the same APs as the UE k. $\mathcal{P}_k$ is adopted to make (3.11) scalable since by definition $\mathcal{P}_k = \{i : \mathbf{D}_k\mathbf{D}_i \neq \mathbf{0}_{LN\times LN}\}$ [27]. Next, the CPUs estimate SINR_k in (2.30) and calculate the SE of the UE k in (2.29).

In the following, it is created the vector $\mathbf{q}_k = [\mathrm{q}_{k1}, ..., \mathrm{q}_{kL}] \in \mathbb{R}^{1\times L}$ to identify the contribution of each AP to the desired signal, where $\mathrm{q}_{kl} = \mathrm{DS}_{kl}$ whether the AP serves the UE (i.e., if $l \in \mathcal{M}_k$) and $\mathrm{q}_{kl} = 0$, otherwise. Then, the elements of $\mathbf{q}_k$ are sorted in descending order leading to the vector $\bar{\mathbf{q}}_k = [\bar{\mathrm{q}}_{k1}, \bar{\mathrm{q}}_{kl'}, ..., \bar{\mathrm{q}}_{kL}]$. The indexes of the APs in the unsorted vector $\mathbf{q}_k$ are stored in the l'-th element of subset $\bar{\mathcal{M}}_k$, where $\bar{L}_k = \left|\bar{\mathcal{M}}_k\right|$. Posteriorly, a cumulative sum is performed, being expressed as

$$\bar{\mathrm{q}}_{kl'}^{\mathrm{sum}} = \begin{cases}\bar{\mathrm{q}}_{kl'} & \text{if } l' = 1\\ \bar{\mathrm{q}}_{kl'} + \bar{\mathrm{q}}_{k(l'-1)}^{\mathrm{sum}} & \text{otherwise}\end{cases}, \tag{3.12}$$

which represents the impact of adding each AP in the desired signal. Finally, one can compute a cumulative SINR as $\mathrm{SINR}_{kl'}^{\mathrm{sum}} = \left|\bar{\mathrm{q}}_{kl'}^{\mathrm{sum}}\right|^2/(\mathrm{IS_k} - \left|\bar{\mathrm{q}}_{kl'}^{\mathrm{sum}}\right|^2 + \sigma_{\mathrm{dl}}^2)$, and calculate $\mathrm{SE}_{kl'}^{\mathrm{sum}}$ as a function of $\mathrm{SINR}_{kl'}^{\mathrm{sum}}$. Therefore, the fine-tuned AP cluster will be found when $\mathrm{SE}_k - \mathrm{SE}_{kl'}^{\mathrm{sum}} \leq \varepsilon$, where ε is the maximum loss limit allowed for the SE.

Then, the CPUs will assign $\mathrm{c}_{kl} = 1$ for the APs in $\bar{\mathcal{M}}_k$, presenting the more substantial contribution to $\mathrm{SE}_{kl'}^{\mathrm{sum}}$ and $\mathrm{c}_{kl} = 0$ for the remaining ones. Recall that we consider a CF system with multiple CPUs. Therefore, each CPU computes $\mathrm{SE}_{kl'}^{\mathrm{sum}}$ and set $\mathrm{c}_{kl} = 0$ or $\mathrm{c}_{kl} = 1$ only for the APs that are linked to it by fronthaul. This fine-tuning scheme is also valid for imperfect knowledge of channel statistics. For this, one should replace $\mathbf{R}_{kl}$ and $\mathbf{R}_{il}$ by $\widehat{\mathbf{R}}_{kl}$ and $\widehat{\mathbf{R}}_{il}$. Algorithm 4 summarizes the entire process.

One can note that only channel statistics are needed to compute the proposed fine-tuning schemes, making them valid for many coherence blocks. In (3.11), the CPUs have to perform about $|\mathcal{P}_k|\,L_iN^3$ complex multiplications per UE, and for scalability purposes, we assume that the CPUs can fine-tune the AP clusters of only LU_{max} UEs, corresponding to the number of connections in the network.

Algorithm 3: Fine-tuning based on power allocation

Input: $\Gamma\%$, $\mathcal{B}_l$;

Sort ϱ_{kl} in descending order for the not mastered UEs
for $k' = 1$ **to** $\left|\bar{\mathcal{B}}_l\right|$ **do**
    Perform a cumulative sum in (3.10)
    Map k' to the unsorted value of k in the subset $\bar{\mathcal{B}}_l$
    if $\varrho^{sum}_{k'l} \leq \Gamma\%$; **then**
        $\mathrm{c}_{kl} = 1$ // Performed in the APs
    else
        $\mathrm{c}_{kl} = 0$
    end
end

Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

Algorithm 4: Fine-tuning based on SE

Input: ε, $\mathcal{M}_k$;

Compute SE_k in (2.29) using (3.11) and create $\mathbf{q}_k$
Sort the elements of $\mathbf{q}_k$ in descending order
for $l' = 1$ **to** $\bar{L}_k$ **do**
    Perform the cumulative sum in (3.12)
    $\mathrm{SINR}^{\mathrm{sum}}_{kl'} = \left|\bar{\mathrm{q}}^{\mathrm{sum}}_{kl'}\right|^2 / (\mathrm{IS_k} - \left|\bar{\mathrm{q}}^{\mathrm{sum}}_{kl'}\right|^2 + \sigma^2_{\mathrm{dl}})$
    Using $\mathrm{SINR}^{\mathrm{sum}}_{kl'}$, compute $\mathrm{SE}^{\mathrm{sum}}_{kl'}$ in (2.29)
    Map l' to the unsorted value of l in the subset $\bar{\mathcal{M}}_k$
    if $\mathrm{SE}_k - \mathrm{SE}^{\mathrm{sum}}_{kl'} \geq \varepsilon$; **then**
        $\mathrm{c}_{kl} = 1$ // Performed in the CPUs
    else
        $\mathrm{c}_{kl} = 0$
    end
end

Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

The fine-tuning scheme presented in Algorithm 4 also works if the CPUs consider the imperfect CSI to calculate DS_k and IS_k. In this case, the closed-form expressions of SE derived for MR in [26] could be employed, but at the cost of $|\mathcal{P}_k|\, L_i(8N^3 - N)$ complex multiplications per UE. Moreover, a pilot assignment strategy has to be performed before the fine-tuning process. It is essential to emphasize that the CPUs assume perfect CSI only to simplify the calculation of DS_k and IS_k. Recall that perfect CSI is not employed in the precoding vectors.

3.3.3 Fine-Tuning Based on Energy Efficiency

The third proposed strategy aims to improve EE. It fine-tunes the AP cluster based on the SE and allocated power ratio ($\mathrm{SE}_k/\varrho_{kl}$). This strategy aims to remove the connections with UEs that achieve high SE and consume a small percentage of the AP power resources. Although this approach may seem counter-intuitive, note that the APs allocate more power (ϱ_{kl}) to the UEs presenting the strongest channel gains in (2.28). Besides, the SE_k increases with the desired signal, which is proportional to the UE channel gain in each serving AP, as (3.11) demonstrates. Thus, if a UE achieves a high SE, receiving only a tiny fraction of power from the AP compared to other UEs, it indicates that this AP is not so fundamental to the SE of this UE. In (2.32), one can also note that the power consumption on fronthaul links ($P_{\mathrm{fh},l}$) is proportional to the SE of the UEs the AP serves in $\mathcal{D}_l$. Therefore, some UEs do not benefit as much from some APs, but contribute to increase $P_{\mathrm{fh},l}$, which decreases the EE in (2.31).

Algorithm 5: Fine-tuning based on EE

Input: ζ, $\mathcal{B}_l$;
for $l = 1$ **to** L **do**
  Compute $\mathrm{SE}_k/\varrho_{kl}$ for all UEs in $\mathcal{B}_l$
  Find the UE with the largest $\mathrm{SE}_k/\varrho_{kl}$, $k_{(l,max)}$
  $\mathrm{r}_{k_{(l,max)}} = \mathrm{SE}_{k_{(l,max)}}/\varrho_{k_{(l,max)}l}$
  for $k'' = 1$ **to** $|\mathcal{B}_l|$ **do**
    Map k'' to the unsorted value of k in subset $\mathcal{B}_l$
    if $(\mathrm{SE}_k/\varrho_{kl})^{-1} > \zeta/\mathrm{r}_{k_{(l,max)}}$ **then**
      $\mathrm{c}_{kl} = 1$ // Performed in the CPUs
    else
      $\mathrm{c}_{kl} = 0$
    end
  end
end
Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

To fine-tune the AP clusters, we proceed as follows: the power allocation is performed in (2.28), and the SE of each UE is computed in (2.29) using (2.30) and (3.11). In the following, the CPUs find the UE presenting the maximum ratio $\mathrm{SE}_k/\varrho_{kl}$ in each AP, such that $k_{(l,max)} = \arg\max_k (\mathrm{SE}_k/\varrho_{kl})$ and $\mathrm{r}_{k_{(l,max)}} = \mathrm{SE}_{k_{(l,max)}}/\varrho_{k_{(l,max)}l}$. For scalability purposes, we adopt a heuristic solution for making the APs drop not mastered UEs (i.e., UEs that are in $\mathcal{B}_l$). We consider that $\mathrm{c}_{kl} = 1$ if $(\mathrm{SE}_k/\varrho_{kl})^{-1} > \zeta/\mathrm{r}_{k_{(l,max)}}$, and $\mathrm{c}_{kl} = 0$ otherwise, where ζ is a project parameter. Each CPU performs these tasks for the APs linked to it by fronthaul. The elements of $\mathcal{B}_l$ are indexed by k'' and Algorithm 5 summarizes the entire process.

4 Reducing Inter-CPU Coordination

The literature considers that UC systems should rely on a network composed of multiple CPUs to be more feasible, as Fig. 8 illustrates. This approach brings the benefits of not overloading a single CPU with the data flows coming from all APs, and some processing tasks, such as data encoding, can be distributed among CPUs. On the other hand, the AP cluster of a specific UE may comprise APs connected to different CPUs, resulting in additional signaling on the backhaul links as the CPUs may have to exchange signals such as DL data to serve distinct AP clusters. In this chapter, the solution proposed in this thesis to control the effects of inter-CPU coordination is presented. The chapter describes the proposed algorithm, which is divided into two stages. The first one associates the UE with a primary CPU, while the second decreases the number of UEs per CPU to mitigate inter-CPU coordination.

4.1 Proposed Algorithm

The proposed method to reduce inter-CPU coordination in UC systems considers that the CPUs and UEs are divided into two classes. Specifically, CPUs are classified as primary and non-primary, while UEs are classified as inter-coordinated and non-inter-coordinated. A primary CPU serves the UE regardless of its channel condition to ensure its connection to the network. In contrast, a non-primary CPU may drop the UE to reduce inter-CPU coordination. Each UE is associated with a primary CPU, but its AP cluster may contain other CPUs (i.e., non-primary ones), as depicted in Fig. 8. Moreover, the UE's primary CPU calls it a non-inter-coordinated UE, while the other CPUs call it an inter-coordinated UE, as Fig. 9 illustrates.

The proposed method aims to reduce the number of inter-coordinated UEs associated with each CPU since inter-CPU coordination results from several UEs being served by multiple CPUs. However, note that the number of non-inter-coordinated UEs on each CPU is not modified since these UEs utilize the CPU as a primary CPU. Therefore, a CPU j does not drop its non-inter-coordinated UEs since CPU j guarantees their connection to the network. For instance, CPU 2 could only drop the UE 2 to reduce the effects of inter-CPU coordination in Fig. 9 since UE 2 is an inter-coordinated UE.

In a nutshell, the proposed strategy is divided into two stages, ant the flowchart exhibited in Fig. 10 provides an overview of the method's operation. In the first one, the new accessing UE k connects with a master AP and a primary CPU to ensure connection to the network, with both the master AP and primary CPU serving the UE regardless of its channel condition. In the following, an arbitrary AP selection scheme is performed,

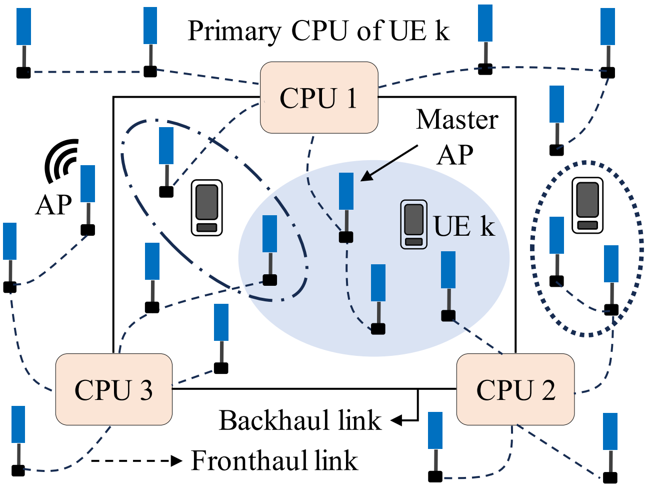


Figure 8 – Illustration of a UC CF massive MIMO network with multiple CPUs. Each CPU coordinates a subset of APs. The CPU 1 is the primary CPU of UE k, while the remaining CPUs are the non-primary CPUs of UE k.

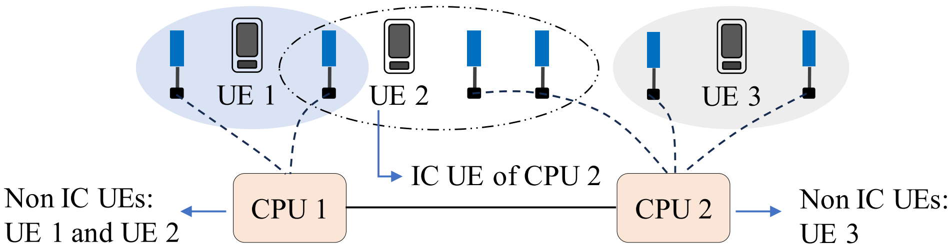


Figure 9 – Illustration of the two classes of UEs. The CPU 1 is the primary CPU of UEs 1 and 2, while CPU 2 is the primary CPU of UE 3 and the non-primary CPU of UE 2. The term IC means inter-coordinated.

and the AP cluster of the UE k is generated. In the second stage, each CPU associated with the AP cluster of the UE k runs a fine-tuning algorithm to decide if they will drop the UE k. The primary CPU keeps the UE k connected, whereas the non-primary CPUs can drop it to reduce inter-CPU coordination.

4.2 First Stage: Master AP and Primary CPU

In the proposed method's first stage, the UE k must associate with a master AP and a primary CPU. To connect with a master AP, the UE solves (3.5). Once a master AP is chosen, it assigns an identifier $\text{ID}_{j_{\text{CPU}}} \in \mathbb{N}$ to UE k, with j being the CPU index. This identifier is a scalar which indicates the CPU that is linked to the master AP by fronthaul. This CPU is considered the primary CPU of the UE k.

In the following, the UE k is associated with a subset of APs ($\mathcal{M}_k$) through an arbitrary AP selection strategy. The connections between the UE and APs are represented

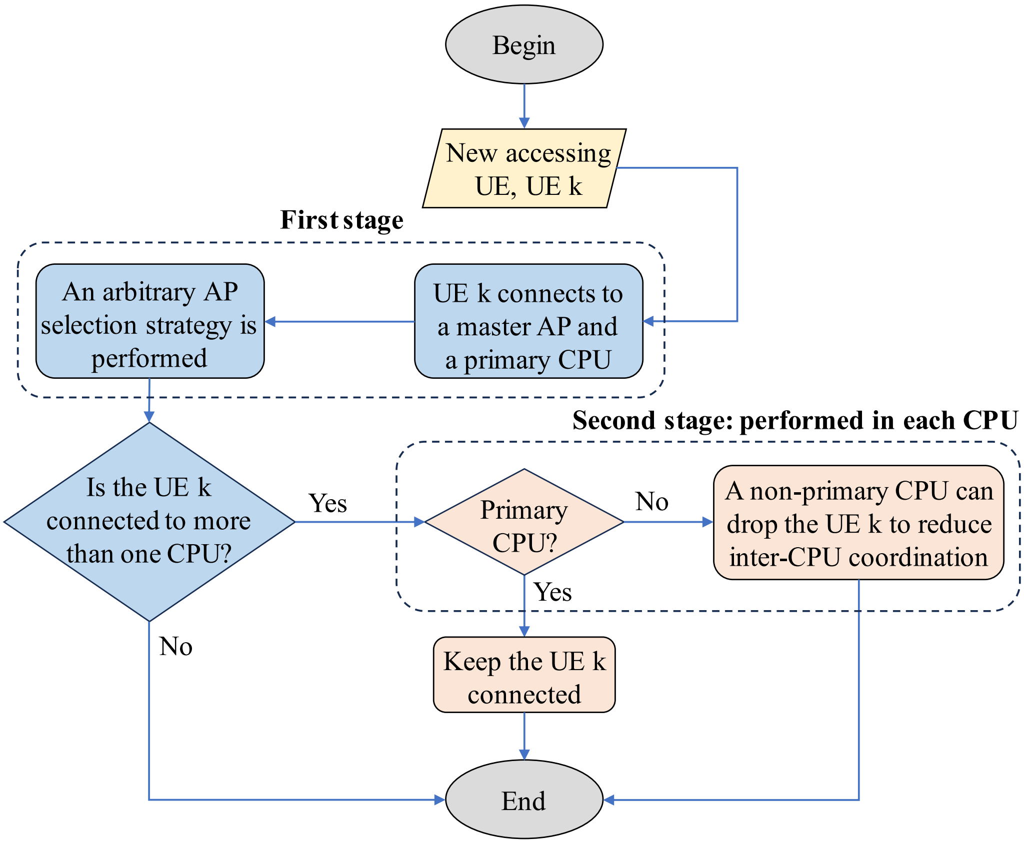


Figure 10 – Simplified flowchart of the proposed strategy.

by vector $\mathbf{e}_k = [\mathrm{e}_{k1}, \ldots, \mathrm{e}_{kL}] \in \mathbb{N}^{1\times L}$, being expressed as

$$\mathrm{e}_{kl} = \begin{cases} 1 & \text{if } l \in \mathcal{M}_k \\ 0 & \text{if } l \notin \mathcal{M}_k \end{cases}. \tag{4.1}$$

The UE k also sends the identifier $\mathrm{ID}_{j_{\mathrm{CPU}}}$ to the APs that are in $\mathcal{M}_k$, and they forward $\mathrm{ID}_{j_{\mathrm{CPU}}}$ to their CPUs, which use $\mathrm{ID}_{j_{\mathrm{CPU}}}$ to identify the primary CPU of UE k. In other words, to check if the UE k is a non-inter-coordinated ($\mathrm{ID}_{j'_{\mathrm{CPU}}} = \mathrm{ID}_{j_{\mathrm{CPU}}}$) UE or a inter-coordinated ($\mathrm{ID}_{j'_{\mathrm{CPU}}} \neq \mathrm{ID}_{j_{\mathrm{CPU}}}$) UE, where $j' \in \{1, \ldots, J\}$.

4.3 Second Stage: Fine-Tuning the AP clusters

In this stage, the CPUs can drop inter-coordinated UEs to mitigate inter-CPU coordination. To this end, this method states that each CPU can serve only a limited number of inter-coordinated UEs, denoted as K_{int}, where $K_{\mathrm{int}} \in \mathbb{N}$ is a system design parameter. Hence, the maximum number of inter-coordinated UEs served by a CPU becomes independent of the number of UEs K. Let $\mathcal{D}_{j'} \subset \{1, \ldots, K\}$ denote the subset of inter-coordinated UEs that have been served by the CPU j'. The fine-tuning procedure

Algorithm 6: Reducing inter-CPU coordination

Input: K_{int}

The UE k connects to a master AP by solving (3.5);

The master AP assigns an identifier $\text{ID}_{j_{\text{CPU}}}$ to the UE k. The CPU j, i.e, the CPU linked to the master AP, will be the primary CPU of UE k;

The UE connects to a subset of APs ($\mathcal{M}_k$) following any AP selection scheme in (2.13). Thus, generating vector $\mathbf{e}_k = [\text{e}_{k1}, \ldots, \text{e}_{kL}]$;

for $j' = 1$ **to** J **do**

  CPU j' computes the number of inter-coordinated UEs it is serving ($|\mathcal{D}_{j'}|$);

  CPU j' identifies its subset of APs serving the UE k ($\mathcal{M}_{kj'}$);

  CPU j' computes the partial sum gain $\text{G}_{kj'} = \sum_{l \in \mathcal{M}_{kj'}} \beta_{kl}$;

  CPU j' identifies its inter-coordinated UE presenting the smallest partial sum gain, i_{min}, where $i_{min} = \arg\min_{i \in \mathcal{D}_{j'}} \text{G}_{ij'}$, with $\text{G}_{ij'} = \sum_{l \in \mathcal{M}_{ij'}} \beta_{il}$;

  if ($\text{ID}_{j'_{\text{CPU}}} = \text{ID}_{j_{\text{CPU}}}$) **or** ($|\mathcal{D}_{j'}| < K_{\text{int}}$) **or** ($\text{G}_{kj'} > \text{G}_{(i_{min})j'}$) **then**

    $\text{f}_{kl} = 1$;

  else

    $\text{f}_{kl} = 0$;

  end

  $\mathbf{c}_k = \mathbf{e}_k \wedge \mathbf{f}_k$;

end

Output: $\mathbf{c}_k = [\text{c}_{k1}, \ldots, \text{c}_{kL}]$.

performed on the CPUs is denoted by vector $\mathbf{f}_k = [\text{f}_{k1}, \ldots, \text{f}_{kL}] \in \mathbb{N}^{1 \times L}$, given by

$$\text{f}_{kl} = \begin{cases} 1 & \text{if } (\text{ID}_{j'_{\text{CPU}}} = \text{ID}_{j_{\text{CPU}}}) \vee (|\mathcal{D}_{j'}| < K_{\text{int}}) \\ 0 & \text{otherwise} \end{cases}, \tag{4.2}$$

for $j' \in \{1, \ldots, J\}$ and $l \in \mathcal{M}_{kj'}$, where $\mathcal{M}_{kj'} \subset \{1, \ldots, L\}$ denotes the subset of APs associated with the UE k that are linked to the CPU j'. Besides, $\vee$ represents the logical operation OR. In (4.2), a non-primary CPU only serves an inter-coordinated UE if it does not reach its maximum capacity (i.e., $|\mathcal{D}_{j'}| < K_{\text{int}}$). To circumvent this issue, (4.2) becomes

$$\text{f}_{kl} = \begin{cases} 1 & \text{if } (\text{ID}_{j'_{\text{CPU}}} = \text{ID}_{j_{\text{CPU}}}) \vee (|\mathcal{D}_{j'}| < K_{\text{int}}) \vee (\text{G}_{kj'} > \text{G}_{(i_{min})j'}) \\ 0 & \text{otherwise} \end{cases}, \tag{4.3}$$

where $\text{G}_{kj'}$ is the partial sum gain, which is computed as $\text{G}_{kj'} = \sum_{l \in \mathcal{M}_{kj'}} \beta_{kl}$. The term partial sum gain means that the CPU j' computes $\text{G}_{kj'}$ only considering the APs that are in $\mathcal{M}_{kj'}$. In (4.3), UE i_{min} represents the inter-coordinated UE served by the CPU j' presenting the smallest partial sum gain, where $i_{min} = \arg\min_{i \in \mathcal{D}_{j'}} \text{G}_{ij'}$, with $\text{G}_{ij'}$ being calculated as $\text{G}_{ij'} = \sum_{l \in \mathcal{M}_{ij'}} \beta_{il}$. In case of $\text{G}_{kj'} > \text{G}_{(i_{min})j'}$, the UE i_{min} is disconnected from all APs linked to the CPU j'. It is worth noting that each CPU operates autonomously and does not exchange any information with other CPUs to compute (4.3). Finally, the final AP cluster of the UE k is given by

$$\mathbf{c}_k = \mathbf{e}_k \wedge \mathbf{f}_k, \tag{4.4}$$

where $\wedge$ is the logical operation AND. From (4.4), one can note that the AP cluster generated in (4.1) is modified by vector $\mathbf{f}_k$. Thus, the final AP cluster of UE k will only be composed of the APs whose CPUs do not drop it. Moreover, the UE k can affect the AP clusters of other UEs since the non-primary CPUs can disconnect the UE i_{min} to serve the UE k in (4.3). Algorithm 6 summarizes the proposed method.

The time complexity of each new UE k to select its coordinating AP by solving (3.5) is $\mathcal{O}(L)$. Assuming the worst case, where all CPUs serve the UE k in (4.1), the time complexity for computing (4.3) and (4.4) is $\mathcal{O}(JK_{\text{int}} \log K_{\text{int}})$, as each CPU has to perform a sort operation for calculating (4.3). Thus, the time complexity of the proposed method is $\mathcal{O}(L + JK_{\text{int}} \log K_{\text{int}})$. It is noteworthy that the time complexity for performing AP selection in (4.1) is not considered because the proposed method is agnostic to the employed AP selection scheme.

5 Scalable User-Centric Cell-Free Massive MIMO with Limited Processing Capacity

In scalable UC systems, the network complexity does not grow with the number of UEs since the number of UEs that each AP serve is limited, i.e., $K_l \leq \tau_p$, where $K_l = |\mathcal{D}_l|$. Therefore, the maximum number of UEs served by each AP remains finite even if the number of UEs K goes to infinity. However, the complexity of performing channel estimation and computing the precoding vectors can still grow with the number of APs, as depicted in Table 2. That is, as L increases, the number of APs connected to the UE k (L_k) can also increase, resulting in more processing complexity from the network, where $L_k = |\mathcal{M}_k|$. This chapter presents the strategy utilized in this thesis to circumvent this issue. Essentially, we rely on a solution where each UE can be associated only with a finite number of APs, denoted as C_{max}, with $L_k \leq \mathrm{C}_{max}$ [52]. We refer to this strategy as maximum AP cluster size control. It is noteworthy that despite having a similar function, the C_{max} on this thesis is fundamentally different from the one presented in [52]. In the proposed approach, C_{max} is a parameter that refers to the system processing capacity limitation that provides a new type of analysis for UC CF massive MIMO systems.

5.1 AP Cluster Size Control

The maximum AP cluster size control procedure is presented in Fig. 11 and can be described as follows: when a new UE k enters the network, it measures the large-scale fading coefficients of the APs in its vicinity, which is calculated according to $\beta_{kl} = \mathrm{tr}\left(\mathbf{R}_{kl}\right)/N$ [26]. Then, it claims a master AP to ensure its connection with at least one AP. For connecting with a master AP, the UE solves (3.5).

After selecting the master AP, the UE k performs any UC AP selection scheme in (2.13). In the following, the CPUs associated with the AP cluster of the UE k share the indexes of the APs serving the UE $(\mathcal{M}_k)$ with each other. Then, the CPUs serving the UE k compute the number of APs serving the UE k, i.e., $L_k = |\mathcal{M}_k|$. If $L_k \leq \mathrm{C}_{max}$, no action is required. Otherwise, the CPUs will drop the connection of the UE k with the E_k APs presenting the weakest channel gains, where E_k denotes the number of APs that exceed C_{max}, which is calculated as $E_k = L_k - \mathrm{C}_{max}$. Let $\mathcal{J}_k$ denote the subset of CPUs associated with the AP cluster of the UE k. The maximum AP cluster size control is performed in J_k CPUs, where $J_k = |\mathcal{J}_k|$.

In order to drop the APs in excess, the J_k CPUs serving the UE k sort the channel gains (β_{kl}) of the APs serving the UE k in ascending order, such that $\tilde{\beta}_{kl'} \leq \cdots \leq \tilde{\beta}_{k(L_k)}$,

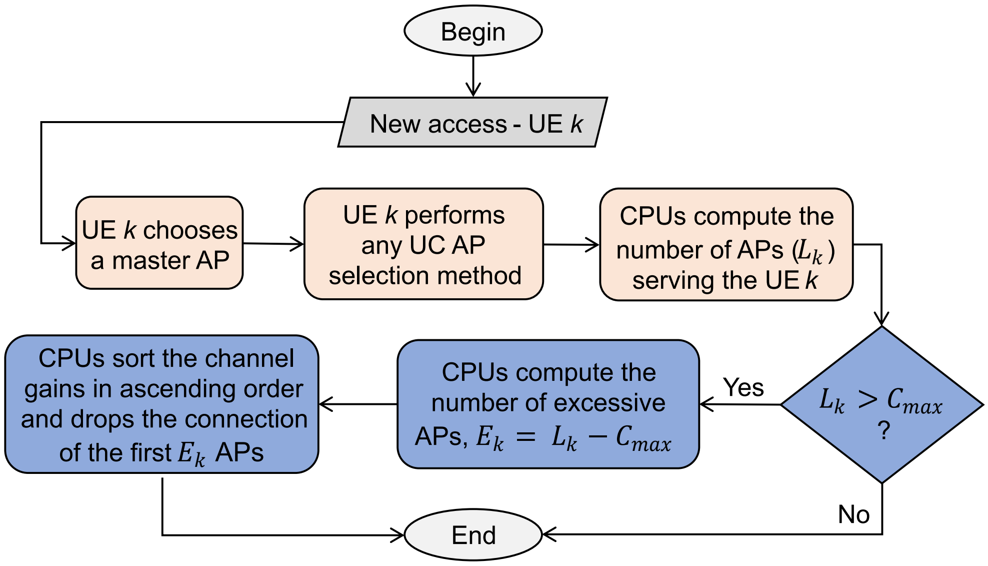


Figure 11 – Flowchart of the AP cluster size control.

where $\tilde{\beta}_{kl'}$ denotes the sorted version of β_{kl}, $\forall l \in \mathcal{M}_k$. The indexes of the APs before the sort operation are stored in the l'-th element of the subset $\bar{\mathcal{M}}_k$. Finally, the CPUs drop the connection of the first E_k APs presenting the smallest channel gains. This procedure can be expressed as

$$\mathrm{c}_{kl} = \begin{cases} 0 & \text{if } l' \leq E_k \\ 1 & \text{otherwise,} \end{cases} \tag{5.1}$$

where l' is mapped to the unsorted value of l in subset $\bar{\mathcal{M}}_k$. Hence, the final AP cluster of UE k will only be composed of the C_{max} APs with the largest channel gains. Algorithm 7 summarizes the maximum AP cluster size control algorithm performed by the CPUs serving the UE k.

Algorithm 7: AP cluster size control

Input: C_{max}

The UE connects to a master AP by solving (3.5) and associates with a subset of APs ($\mathcal{M}_k$) in (2.13);

Identify the J_k CPUs serving the UE; // $J_k = |\mathcal{J}_k|$

// The J_k CPUs perform AP cluster size control:

if $L_k > \mathrm{C}_{max}$ **then**

  $E_k = L_k - \mathrm{C}_{max}$; // where $L_k = |\mathcal{M}_k|$

  Sort the channel gains of the APs serving the UE in ascending order, such that $\tilde{\beta}_{kl'} \leq \cdots \leq \tilde{\beta}_{k(L_k)}$;

  for $l' = 1$ **to** E_k **do**

    Map l' to the unsorted value of l in subset $\bar{\mathcal{M}}_k$;

    $\mathrm{c}_{kl} = 0$; // Computed in (5.1)

  end

end

Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

The time complexity of the proposed technique is calculated as follows: the complexity for selecting a master AP by solving (3.5) is $\mathcal{O}(L)$. The time complexity to perform AP cluster size control in each CPU j is $\mathcal{O}\left(\left|\mathcal{K}_j^{\text{all}}\right| \log \left|\mathcal{K}_j^{\text{all}}\right|\right)$, since each CPU has to perform a sort operation before computing (5.1). Therefore, the overall time complexity can be expressed as $\mathcal{O}\left(L + \sum_{j=1}^{J} \left|\mathcal{K}_j^{\text{all}}\right| \log \left|\mathcal{K}_j^{\text{all}}\right|\right)$.

5.2 AP Cluster Adjustment

In this section, a heuristic method that adjusts the AP clusters according to the network implementation is proposed. Such method holds for any UC AP selection strategy, i.e., with and without processing capacity limitation. Moreover, it is a heuristic technique because only heuristic solutions are scalable [26]. In a nutshell, the UEs are associated with a subset of APs following any AP selection process. Then, the proposed method aims to simultaneously reduce the number of UEs served by each AP l (K_l) and the number of APs connected to each UE k (L_k) while keeping the SE under minor degradation. In this context, it is a novel way to reduce the CC and increase EE in scalable UC CF massive MIMO systems. The analysis also assumes that each UE connects to a master AP by solving (3.5).

5.2.1 AP Cluster Adjustment in the Distributed Implementation

In the distributed implementation, the proposed technique exploits the local long-term CSI at each AP and intends to reduce K_l without causing significant SE degradation. When all APs are involved, the average value of L_k is also reduced. It is noteworthy that L_k is not directly decreased in distributed implementation, and neither could it be since it would require global long-term CSI at each AP.

The adjustment of the AP cluster relies on two proposed metrics: (i) the partial channel strength indicator ($\bar{\beta}_{kl}$) and (ii) the total channel strength indicator ($\bar{\beta}_l$). We utilize these metrics to prevent the less fortunate UEs from being easily dropped by the AP. Thus, they do not directly represent the long-term CSI of the UEs that the AP serves. Instead, they are the long-term CSI raised to a normalization exponent, defined as λ_l, which provides a better balance between the channel gains of the most and less fortunate UEs served by the AP, such that $0 < \lambda_l < 1$. Without this normalization, the AP could easily drop a UE presenting a weaker channel gain if the AP was also serving UEs with stronger channel gains. Nevertheless, these differences can be decreased when the channel gains are raised to a power lower than one and greater than zero, such as λ_l. The partial channel strength indicator is given by $\bar{\beta}_{kl} = (\beta_{kl})^{\lambda_l}$, where

$$\lambda_l = \frac{\min_{k \in \mathcal{D}_l}(\beta_{kl})}{\max_{k \in \mathcal{D}_l}(\beta_{kl})}. \tag{5.2}$$

The second metric, called total channel strength indicator, is calculated as $\bar{\beta}_l = \sum_{k \in \mathcal{D}_l} \bar{\beta}_{kl}$. In the proposed method, the two metrics are used by each AP l to compute

$$\bar{\beta}_{l,-k} = \bar{\beta}_l - \bar{\beta}_{kl}, \tag{5.3}$$

$\forall k \in \mathcal{D}_l$. The purpose of calculating $\bar{\beta}_{l,-k}$ is to evaluate how much the total channel strength indicator $\bar{\beta}_l$ is reduced by dropping the UE k from the AP l. After computing (5.3), the AP keeps the connection of UE k, only if

$$\mathrm{c}_{kl} = \begin{cases} 1 & \text{if } (\text{UE } k \in \mathcal{A}_l) \vee \left(\bar{\beta}_{l,-k} \leq \bar{\beta}_l^{mean}\right) \\ 0 & \text{otherwise,} \end{cases} \tag{5.4}$$

where $\bar{\beta}_l^{mean} = \sum_{k \in \mathcal{D}_l} \bar{\beta}_{l,-k}/K_l$ is a threshold value and $\mathcal{A}_l \subset \mathcal{D}_l$ is the subset of UEs that AP l serves as a master AP. One can note that the term $\bar{\beta}_{l,-k}$ has to be smaller than $\bar{\beta}_l^{mean}$, because $\bar{\beta}_{l,-k}$ will be small if the UE k has a large partial channel strength indicator $\bar{\beta}_{kl}$, since $\bar{\beta}_{l,-k} = \bar{\beta}_l - \bar{\beta}_{kl}$. Meanwhile, $\bar{\beta}_{l,-k}$ will be large if the UE k adds only a marginal gain to the total channel strength indicator $\bar{\beta}_l$. That is, if $\bar{\beta}_{kl}$ represents a considerable percentage of $\bar{\beta}_l = \sum_{k \in \mathcal{D}_l} \bar{\beta}_{kl}$, the term $\bar{\beta}_l$ will be significantly reduced if the UE k is disconnected from AP l. Algorithm 8 summarizes the AP cluster adjustment in the distributed implementation.

Algorithm 8: AP cluster adjustment in the distributed implementation

Input: $k' = 1, ..., K_l$

Compute λ_l and $\bar{\beta}_{kl} = (\beta_{kl})^{\lambda_l}$, $\forall k \in \mathcal{D}_l$, // Partial channel strength indicator calculated in the APs

$\bar{\beta}_l = \sum_{k \in \mathcal{D}_l} \bar{\beta}_{kl}$;// Total channel strength indicator

Compute $\bar{\beta}_{l,-k}$ and $\bar{\beta}_l^{mean} = \sum_{k \in \mathcal{D}_l} \bar{\beta}_{l,-k}/K_l$.

for $k' = 1$ **to** K_l **do**

  Map k' to the value of k in subset $\mathcal{D}_l$.

  if $k \notin \mathcal{A}_l$ **and** $\bar{\beta}_{l,-k} \geq \bar{\beta}_l^{mean}$ **then**

    $\mathrm{c}_{kl} = 0$; // Remove the UE from the AP

  end

end

Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

5.2.2 AP Cluster Adjustment in the Centralized Implementation

In the centralized implementation, the long-term CSI of APs and UEs is available at the CPUs [26,28]. Hence, the proposed method exploits the global long-term CSI to reduce L_k. At first, reducing L_k may appear counter-intuitive since the centralized implementation has a better interference suppression capability. However, since CC grows with the number of APs serving the UE (recall that $L_k = |\mathcal{M}_k|$), the AP cluster expansion will not always be beneficial, and reducing L_k may be necessary even in this implementation. In the centralized implementation, the AP cluster adjustment is also performed by the J_k CPUs associated with the AP cluster of the UE k, which are denoted as $\mathcal{J}_k$, where $J_k = |\mathcal{J}_k|$. Moreover, the J_k CPUs need to share the indexes of the APs serving the UE ($\mathcal{M}_k$) with each other, as in Section 5.1.

The partial channel strength indicator is now calculated in the CPUs as $\bar{\beta}_{kl} = (\beta_{kl})^{\lambda_k}$, where λ_k introduces a balance between the serving APs presenting the smallest and highest channel gain to the UE k. The CPUs compute λ_k as

$$\lambda_k = \frac{\min_{l\in\mathcal{M}_k}(\beta_{kl})}{\max_{l\in\mathcal{M}_k}(\beta_{kl})}. \tag{5.5}$$

The total channel strength indicator is computed as $\bar{\beta}_k = \sum_{l\in\mathcal{M}_k}\bar{\beta}_{kl}$. Then, the CPUs calculates the contribution that each AP brings to $\bar{\beta}_k$ as

$$\bar{\beta}_{k,-l} = \bar{\beta}_k - \bar{\beta}_{kl}, \tag{5.6}$$

$\forall l \in \mathcal{M}_k$. Therefore, the CPU connected to the AP l keeps the connection of AP l with the UE k only if

$$\mathrm{c}_{kl} = \begin{cases} 1 & \text{if } (\text{UE } k \in \mathcal{A}_l) \vee \left(\bar{\beta}_{k,-l} \leq \bar{\beta}_k^{mean}\right) \\ 0 & \text{otherwise,} \end{cases} \tag{5.7}$$

where $\bar{\beta}_k^{mean} = \sigma_{si}/2 + \sum_{l\in\mathcal{M}_k}\bar{\beta}_{k,-l}/L_k$, with σ_{si} denoting the standard deviation of $\bar{\beta}_{k,-l}$, $\forall l \in \mathcal{M}_k$. The term σ_{si} is utilized to make the CPUs drop fewer APs from the AP cluster of UE k to exploit the centralized implementation's capacity in improving SE. It is worth noting that only the CPUs associated with the AP cluster of the UE run the proposed method. Algorithm 9 summarizes the AP cluster adjustment in the centralized implementation.

Algorithm 9: AP cluster adjustment in centralized implementation

---

Input: $l' = 1, ..., L_k$, $\mathcal{J}_k$

Compute λ_k and $\bar{\beta}_{kl} = (\beta_{kl})^{\lambda_k}$, $\forall l \in \mathcal{M}_k$ // Partial channel strength indicator calculated in the CPUs

$\bar{\beta}_k = \sum_{l\in\mathcal{M}_k}\bar{\beta}_{kl}$ // Total channel strength indicator

Compute $\bar{\beta}_{k,-l}$ and $\bar{\beta}_k^{mean} = \sigma_{si}/2 + \sum_{l\in\mathcal{M}_k}\bar{\beta}_{k,-l}/L_k$

for $l' = 1$ **to** L_k **do**

  Map l' to the value of l in subset $\mathcal{M}_k$

  if $k \notin \mathcal{A}_l$ **and** $\bar{\beta}_{l,-k} \geq \bar{\beta}_l^{mean}$ **then**

    $\mathrm{c}_{kl} = 0$

  end

end

Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

5.2.3 Pros and Cons of the two AP Clusters Adjustments

The utilization of the proposed method on a distributed implementation enables a fronthaul signaling reduction since the number of data flows on the fronthaul is proportional to K_l in Table 1. Besides, it allows the AP to carry out fewer operations while attaining the same SE performance, increasing the system's EE. Utilizing the proposed method in a centralized implementation also allows significant savings in CC resources. Nonetheless, it does not directly reduce the number of data flows in the fronthaul links since the data traffic is not proportional to K_l, but to the number of antennas per AP, N. It is worth noting that this thesis has considered that the AP cluster adjustment is only activated when λ_l and λ_k are lesser than a threshold Θ to avoid excessive adjustments, where Θ is a project parameter. We have set $\Theta = 10^{-2}$ and $\Theta = 10^{-3}$ for the distributed and centralized implementations, respectively.

6 Numerical Results

This chapter details the numerical simulations and results carried out in this thesis. The results are divided in three sections and are organized as follows: Section 6.1 introduces the numerical results associated with the matched-decision AP selection framework and the fine-tuning algorithms, described in Chapter 3. Section 6.2 exhibits the simulation results related to the proposed method to reduce the effects of inter-CPU coordination, discussed in Chapter 4. Section 6.3 details our findings for the strategies presented in Chapter 5, which consist in limiting the processing capacity of UC systems and adjust the AP clusters of UEs according to each network implementation (i.e., centralized or distributed processing). A description of the simulation parameters and assumptions employed in the results are given bellow.

We consider a CF network consisting of L APs, each equipped with N antennas. Each AP can serve up to U_{max} UEs, which describes a processing capability limitation of the AP and allows the system to achieve scalability. The K UEs are uniformly distributed over a square area of 1×1 km, and the distribution of the APs follows a hard core point process (HCPP)[1]. After the APs positioning, the coverage area is divided into J rectangle regions of the same size, each consisting of a CPU coordinating approximately L/J APs, where we have set $J = 4$. The values of L, N, U_{max}, and K vary and are specified throughout the results. In order to provide a better balance as to the amount of interference that affects each AP, we employ the wrap-around technique [4]. We focus on DL channels and consider $\tau_c = 200$ samples in each coherence block. The pre-log factor is set to $P_f = \tau_d/\tau_c$ for perfect knowledge of channel statistics, where $\tau_p = 10$, and $\tau_d = 190$. For imperfect knowledge, $P_f = 1 - (\tau_p/\tau_c) - \alpha$, where $\alpha = N_R \tau_p K/\tau_s \tau_c$. We consider that $B_s = 100$ MHz and $T_s = 0.5$ s, such that $\tau_s = 250000$ [92]. To calculate α and perform the correlation matrix estimation, we assume $N_R = 400$ and $N_{\mathbf{\Psi}} = 800$.

For computing the centralized power allocation (ϱ_k) we consider the following fractional power parameters $\upsilon' = -0.5$ and $\kappa' = 0.5$ [27]. In the UL direction, we assume that each UE transmits the pilot signals with full power [26]. The parameters for EE are set as $\nu_l = 0.4$, $P_{\text{tc},l} = 0.2$ W, $P_{0,l} = 0.825$ W, and $P_{\text{ft},l} = 0.25$ W/(Gbit/s) [24]. We perform Monte-Carlo simulations to evaluate the system's performance in terms of average and cumulative distribution function (CDF) of the SE. We also evaluate the average numbers of APs connected to each UE (L_k) and UEs per AP (K_l). The propagation model

[1] We use a HCPP because it adds a better spacing regularity between the APs that would not be possible in a uniform distribution. In this method, the distance between any two APs cannot be smaller than $d_{\text{min}} = \sqrt{A/L}$, where A is the coverage area in square meters. The first step is to randomly drop the APs based on a homogeneous Poisson point process with mean a rate $1/d_{\text{min}}$, then randomly update the location of APs that do not meet the spacing requirement until it is fulfilled.

adopted is in accordance with the 3GPP LoS/NLoS Urban Micro (UMi) path-loss model defined in the Technical Report (TR) 38.901 [98]. We estimate the perfect correlation matrices of NLoS channels $\mathbf{R}_{kl}^{\text{NLoS}}$ according to the local scattering spatial correlation model presented in [4, Sec. 2.6]. The parameter values used to set the entries for the UMi model and $\mathbf{R}_{kl}^{\text{NLoS}}$ can be found in Table 3.

Table 3 – Parameters assumed for the UMi path-loss and local scattering spatial correlation model utilized to compute the correlation matrix of NLoS channels.

Parameter	Value
Effective environment height, h_E	1.0 m
Shadow fading standard deviation, σ_{SE}	4 dB
Antenna height AP, UE - h_{AP}, h_{UE}	11.65 m, 1.65 m
Rx noise figure (NF)	8 dB
Frequency center, bandwidth (B_s)	3.5 GHz, 100 MHz
angular standard deviations (ASDs)	$\sigma_\varphi = \sigma_\theta = 15°$
Antenna spacing	1/2 wavelength distance

Throughout the results, we modify the scalable CF AP selection method to consider the restriction U_{max}. For pilot allocation, we consider the algorithm presented in [26]. This is because we intend to compare the proposed solutions with baseline AP selection schemes, such as the scalable CF scheme, which relies on this pilot assignment strategy to generate the AP clusters. This method assumes that the pilot assignment is performed by the AP with the strongest channel gain in the AP cluster of each UE. Moreover, the pilot assigned to each UE is the one that causes the least pilot contamination in the master AP. The literature has introduced several strategies for pilot assignment in massive MIMO theory [99–101]. One can investigate how different AP selection schemes are affected by distinct pilot assignment strategies and vice-versa. However, a more profound discussion involving pilot allocation strategies is out of the scope of this thesis.

We compare the achievable SE results for different precoding choices. Recall that the MMSE precoding is a signal processing technique that maximizes the SINR for all UEs in the network based on channel estimates available on the CPU, efficiently suppressing interference among them. The distributed version of MMSE is called L-MMSE precoding, which uses the local channel estimates available at each AP. These methods are not scalable since they need channel estimates of all UEs, but one can modify them to fulfill the scalability requirements. The key features of each are summarized below [27]:

- Distributed implementation
    - MR scheme: low-complexity precoding that maximizes the ratio between the received power and the square norm of the UL combing vector but cannot efficiently mitigate interference among UEs.

  - LP-MMSE scheme: an adaptation of the L-MMSE precoding that suppresses interference only from UEs served by the AP using locally available channel estimates, that is, with no cooperation among APs to this end. This precoding has higher complexity than MR method but can provide better SE.

- Centralized implementation

  - P-MMSE scheme: the only difference between this precoding and the MMSE is that it does not consider all UEs in the UL combining vector, but only the subset of UEs partially served by the same APs. The P-MMSE better mitigates interference than LP-MMSE, as the CPU entity has access to the channel estimates of several UEs, not only using local estimations.
  - P-RZF scheme: this technique simplifies the P-MMSE precoding by neglecting the estimation errors correlation matrix, allowing it to reduce time complexity. However, it still can suppress interference of the subset of UEs partially served by the same APs. It also generally performs better than LP-MMSE.

6.1 Simulation Results: Chapter 3

This section presents the numerical results associated with the matched-decision AP selection framework and the fine-tuning algorithms described in Chapter 3. This section follows the same parameters and assumptions discussed previously but with some particularities, since it relies on specific AP selection methods. These specificities are: the total transmission powers of the UEs and APs are $\eta_i = 100\,\text{mW}$, $\varrho_l = 1\,\text{W}$, respectively. The threshold value of the LSFB algorithm is set to $\delta\% = 99.9$ and it is considered that $\gamma = -50\,\text{dB}$ in the proposed method. Moreover, we set $\Gamma\% = 98$, $\varepsilon = 0.02$ and $\zeta = \tau_p$ for the fine-tuning schemes. The proposed solutions are compared with four other AP selection algorithms described earlier in Section 3.2, which are the canonical CF [23], UCC [25], LSFB [24], and scalable CF [26]. Besides, it is considered that the channel vector between a UE k and AP l undergoes an independent correlated Rayleigh fading, such as in (2.5).

6.1.1 Cumulative Distribution Function

We start by evaluating the CDF of the achievable DL SE in a network consisting of $L = 100$ APs and $K = 20$ UEs. In Fig. 12, we compare the performance of the AP selection methods considering two scenarios and assuming perfect knowledge of channel statistics. The results compare APs that can serve a lower number of UEs ($U_{max} = 4$) with APs that can deal with more UEs ($U_{max} = 10$), which represent different processing capacity of their hardware.

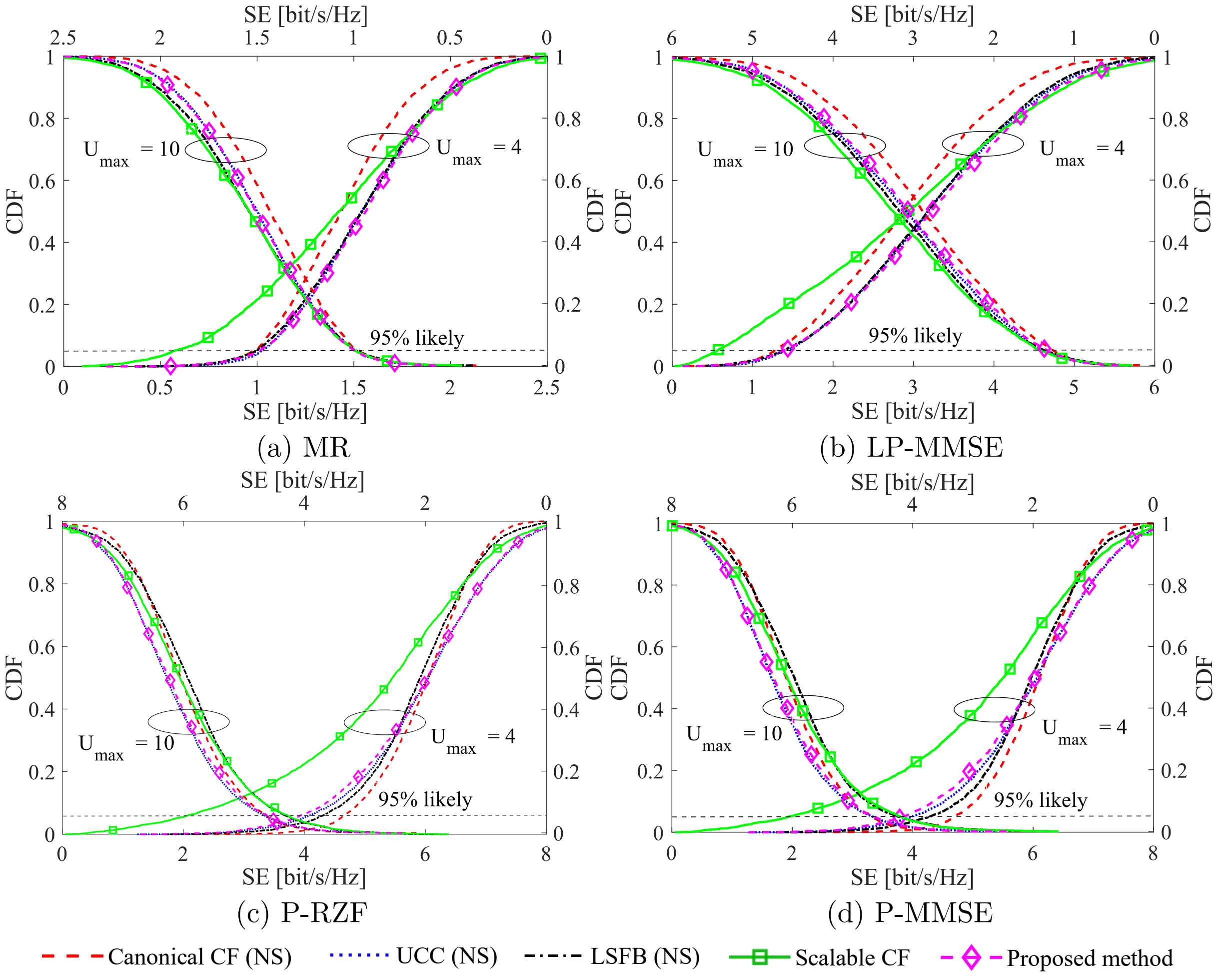


Figure 12 – Comparison of DL SE per UE of the proposed AP selection method with canonical CF [23], UCC [25], LSFB [24] and scalable CF [26]. Parameters setting: $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

In Fig. 12, the proposed solution can outperform the SE's of the remaining methods for the 95% likely UEs, when $U_{max} = 4$. It increases the SE of the 95% likely UEs by approximately 100% both for the P-MMSE and P-RZF compared to the scalable CF scheme, where the P-RZF achieves the same SE of the P-MMSE even though it is a less complex technique. In the distributed implementation, we observe a gain of about 163% using LP-MMSE and 75% with MR compared with [26], which corresponds to the expectations of both methods regarding interference mitigation. Additionally, the proposed method provides higher SEs with approximately the same number of APs connected per UE for $U_{max} = 4$, according to Table 4. It is similar to UCC but with the advantage of being scalable and presenting less time complexity.

In Fig. 12, the increase in SE for the 95% likely UEs for $U_{max} = 4$ is related to the final AP cluster stage, which makes the worst UEs to connect to more APs. On the other hand, the 50% and 10% likely UEs benefit less from this stage, even if their SEs raise more. For example, an increase of 1 (bit/s/Hz) can substantially enhance the SE of the 95% likely UEs in percentage, while 1.2 (bit/s/Hz) will not present the same impact for the

50% and 10% likely UEs. Moreover, the intermediate AP cluster was crucial to improve performance. One can note that even though our solution makes the AP clusters comprise more APs, the average number of APs connected to each UE is not that different in Table 4 compared to the scalable CF method, since $U_{max} = 4$ is relatively small. Therefore, the matched-decision strategy enabled UEs and APs to establish more suitable connections, which helped in enhancing SE.

Table 4 – Mean value and standard deviation (STD) of the number of APs connected to each UE for different AP selection methods.

AP selection method	$U_{max} = 4$		$U_{max} = 10$		Complexity
	Mean	STD	Mean	STD	
UCC (NS) [25]	20	0	50	0	$\mathcal{O}(LK \log K)$
LSFB (NS) [24]	30.28	1.46	30.28	1.46	$\mathcal{O}(L \log L)$
Scalable CF [26]	18.57	0.93	24.67	2.69	$\mathcal{O}(L\tau_p)$
Proposed method	19.94	0	49.94	0.57	$\mathcal{O}(LU_{max})$

In Fig. 12, the gains offered by our method are not that impressive for $U_{max} = 10$, and some baseline solutions were slightly better in distributed implementation. As the APs can serve more UEs when $U_{max} = 10$, our solution makes the AP clusters even larger (as Table 4 shows), which generates more interference in the DL direction. Therefore, the proposed scheme has to be used jointly with more robust precoding techniques, such as P-MMSE and P-RZF, to provide gains in SE when $U_{max} = 10$. Otherwise, baseline solutions can present higher SEs since they serve the UEs using fewer APs, as Table 4 demonstrates. Nonetheless, the proposed method can present the lowest time complexity for $U_{max} < \tau_p$.

In Fig. 12, one can note that the proposed method allows the systems that employ APs with small U_{max} to provide SEs as high as those with a higher U_{max}. For instance, for the P-MMSE, the SE of the 50% likely UEs is about 6 (bit/s/Hz) and 6.3 (bit/s/Hz) for $U_{max} = 4$ and $U_{max} = 10$, respectively. These insights reveal that a network that employs APs serving many UEs does not necessarily provide a higher capacity. Furthermore, Fig. 12 also indicates that even if an AP can serve more UEs, it could reduce U_{max} through software to improve SE in some scenarios. Such results may also inspire future publications involving scalable UC networks by showing that U_{max} must not be too small or too large but properly suited to the network conditions. For instance, almost all AP selection schemes can outperform the canonical CF (at least in these controlled simulations) in Figs. 12 (a) and (b), emphasizing that an AP serving a higher number of UEs do not necessarily bring the highest SEs, as the more UEs the AP serves, the less is the allocated power and the higher is the interference.

6.1.2 Similarities with other AP Selection Methods

Recall that the proposed method can also provide scalability for other AP selection schemes. Fig. 13 compares the LSFB (NS) strategy with the MD LSFB described in Section 3.2. Note that the MD LSFB strategy can achieve similar SEs as the LSFB with the advantage of being scalable. Moreover, it can perform as great as the scalable CF scheme. It can be noted in Fig. 12 that the LSFB (NS) scheme performs better than the scalable CF scheme for $U_{max} = 4$ and matches this one when $U_{max} = 10$. As the scalable version of the LSFB does not present notable performance losses in Fig. 13, one can conclude that the MD LSFB presents SE levels as high as the scalable CF scheme.

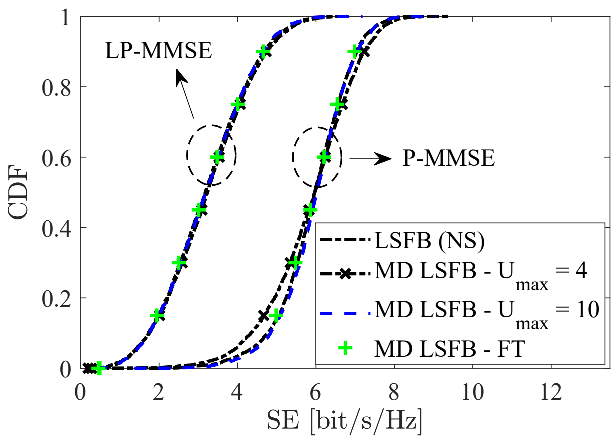
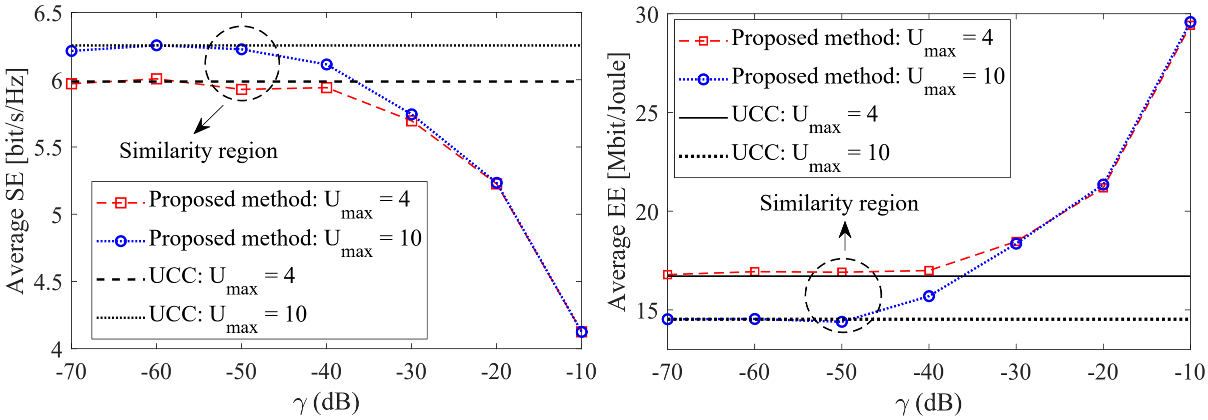


Figure 13 – Comparison of DL SE achieved when using LSFB and its scalable version, the MD LSFB. It is also presented a fine-tuned version of the MD LSFB (based on Algorithm 4) for $U_{max} = 10$, the MD LSFB-FT. Parameters setting: $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

A similar result can be observed for the UCC scheme since it is the method that most closely matches our scheme in Fig. 12. As Fig. 14 shows, the similarities between the proposed method and the UCC are related to the value of γ. For $\gamma < -40$ dB, the UE discovers a large number of APs in its vicinity, which makes the decisions of the APs predominant. Therefore, if all UEs choose many APs to connect, the APs will select those with the stronger channel gain in the intermediate stage. Thus, for $\gamma < -40$ dB, the proposed scheme is probably approaching a scalable version of UCC, which is called MD UCC. However, as γ increases, the UE decisions become even more restricted in the intermediate stage, i.e., the UE selects fewer APs in (3.6) and the similarities disappear. One can note that the similarity region (i.e., $\gamma < -40$ dB) allows the proposed method to reach high SEs while reducing EE. Therefore, the matched decision scheme must operate outside the similarity region to improve EE. That is, $\gamma > -40$ dB.

6.1.3 Performance of Fine-Tuning AP selection

Regarding the fine-tuning AP selection methods of Algorithms 3 and 4, one can observe in Table 5 that they substantially reduce the number of APs connected to each



(a) SE:P-MMSE (b) EE:LP-MMSE

Figure 14 – Average DL SE and EE achieved by the proposed solution for different values of γ. Parameters setting: $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

UE (especially for $U_{max} = 10$) while achieving similar results in terms of SE. These results demonstrate that fine-tuning AP selection schemes can potentiate the network performance, as the number of complex multiplications to precoding signals and estimate channels is proportional to L_k and K_l. In Table 5, one can note that the fine-tuning based on power allocation (Algorithm 3) can improve SE in LP-MMSE while reducing it in P-MMSE. As the distributed nature of this strategy aims to reduce the number of UEs connected in each AP (K_l), it helps the interference mitigation of LP-MMSE. Therefore, it is more recommendable for distributed implementation.

On the other hand, both P-MMSE and LP-MMSE benefit from the fine-tuning based on SE (Algorithm 4) since this method reduces the number of APs connected to each UE by looking to the UE side. That is, it tries to keep the SE of the UE under minor degradation, while Algorithm 3 only considers the allocated power that each UE receives from the AP. It is worth noting that these fine-tuning methods work in any AP selection scheme. For instance, Fig. 13 shows that we can keep the SE of the MD LSFB strategy almost unchanged while reducing the average number of APs per UE (L_k) from 30.38 to 26.51 and the average number of UEs per AP (K_l) from 6.05 to 5.3. In Fig. 13, the fine-tuning strategy based on SE is employed for $U_{max} = 10$, and the fine-tuned version of MD LSFB is called MD LSFB - FT.

In Fig. 15, we analyze the impacts of Algorithm 5 on the system's EE. Specifically, it is evaluated the EE of the MD LSFB scheme vs. the variation of ζ. The more ζ grows, the more the APs disconnect UEs (i.e., K_l reduces). Therefore, the fine-tuning of Algorithm 5 is not activated when $\zeta = 0$. When ζ is maximum, the number of UEs the AP serves assumes the lowest value. It is considered that $0 \leq \zeta \leq \tau_p$, where $\tau_p = 10$. The results indicate that we can improve the EE up to 43.3% in the LP-MMSE for $U_{max} = 4$. These values are achieved by comparing the EE values in $\zeta = 0$ and $\zeta = 10$.

Table 5 – Average number of APs per UE (L_k), UEs per AP (K_l) and DL SE for the proposed AP selection method before and after applying the fine-tuning methods from Algorithms 3 to 5. Parameters setting: $L = 100$, $N = 1$ and $K = 20$, $\Gamma\% = 98$, $\varepsilon = 0.02$, and $\zeta = \tau_p$. Perfect knowledge of channel statistics.

AP selection method	$U_{max} = 4$			
	UEs per AP	APs per UE	LP-MMSE SE	P-MMSE SE
Proposed method	3.98	19.94	3.26 bit/s/Hz	5.92 bit/s/Hz
Algorithm 3	2.78	13.9	3.31 bit/s/Hz	5.57 bit/s/Hz
Algorithm 4	3.19	15.96	3.28 bit/s/Hz	5.89 bit/s/Hz
Algorithm 5	1.45	7.28	3.34 bit/s/Hz	4.77 bit/s/Hz
AP selection method	$U_{max} = 10$			
	UEs per AP	APs per UE	LP-MMSE SE	P-MMSE SE
Proposed method	9.98	49.94	3.09 bit/s/Hz	6.22 bit/s/Hz
Algorithm 3	7.55	37.78	3.15 bit/s/Hz	6.06 bit/s/Hz
Algorithm 4	7.76	38.8	3.14 bit/s/Hz	6.18 bit/s/Hz
Algorithm 5	2.38	11.9	3.32 bit/s/Hz	5.24 bit/s/Hz

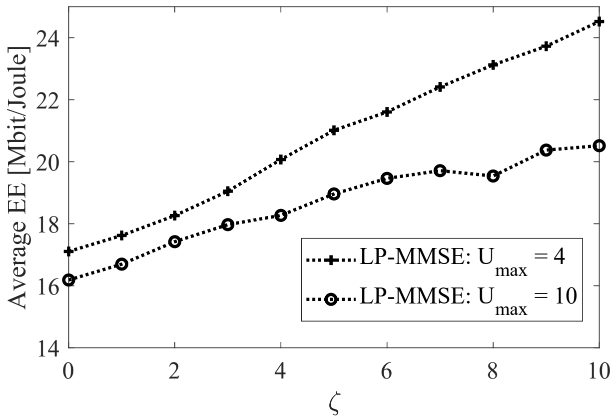


Figure 15 – Average EE achieved by the MD LSFB scheme after fine-tuning the AP clusters based on EE (Algorithm 5). Parameters setting: $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

The fine-tuning scheme of Algorithm 5 increases the EE by reducing the average number of APs per each UE (L_k) from 30.38 to 8.7043 for $U_{max} = 10$ and from 19.12 to 4.38 for $U_{max} = 4$, when the MD LSFB scheme is employed. Moreover, it also reduces the average number of UEs per AP (K_l). It decreases from 6 to 1.74 when $U_{max} = 10$ and from 3.81 to 0.87 for $U_{max} = 4$. These low values indicate that the Algorithm 5 turned off several APs to improve EE. Nevertheless, Algorithm 5 can decrease the SE in the P-MMSE scheme, as Table 5 illustrates. On the other hand, Algorithm 5 can slightly increase the SE of the LP-MMSE as depicted in Table 5. Therefore, it is more suitable for a distributed implementation. Similar results are observed in the other AP selection methods. For instance, the EE can increase up to 40% for $U_{max} = 10$ in the LP-MMSE precoding when the scalable CF scheme is employed.

6.1.4 Impacts of Imperfect Knowledge of Channel Statistics

The impacts of imperfect knowledge of the correlation matrices are illustrated in Fig. 16, with R representing the correlation matrices. It is possible to observe that the scalable CF scheme is the most affected by the imperfect $\widehat{\boldsymbol{\Psi}}_{tl}$ and $\widehat{\mathbf{R}}_{kl}$ when $U_{max} = 4$. In this scenario, the proposed method can offer gains of up to 315 % in the SE of the 95 % likely UEs compared to the scalable CF strategy. Nonetheless, both strategies are not greatly affected when $U_{max} = 10$.

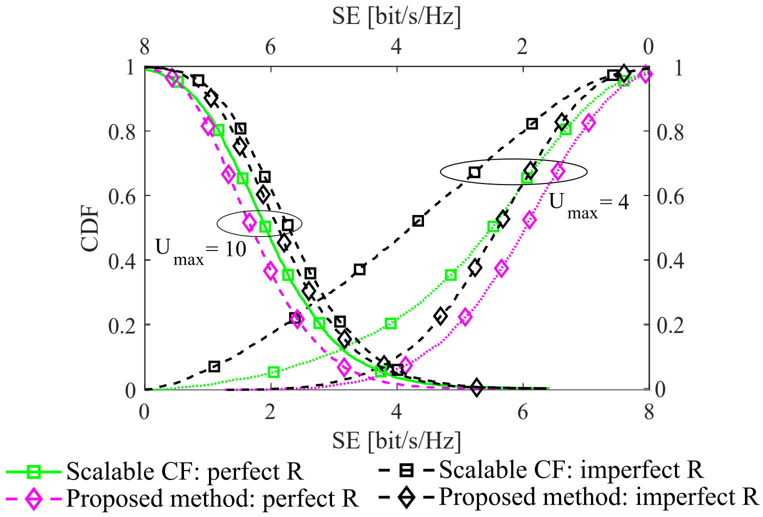


Figure 16 – Comparison of SE considering imperfect knowledge of channel statistics in the P-MMSE precoding. Parameters setting: $L = 100$, $N = 1$, and $K = 20$. The letter R represents the correlation matrices.

The evaluation of imperfect knowledge of channel statistics was carried out for all AP selection schemes on all precoding vectors previously described. However, we showed only the results of P-MMSE precoding in two AP selection schemes to avoid redundancies. In general, although not shown in the figures, all AP selection methods presented only small performance losses when $\widehat{\boldsymbol{\Psi}}_{tl}$ and $\widehat{\mathbf{R}}_{kl}$ are imperfect. The most degraded one was the scalable CF for all precoding techniques when $U_{max} = 4$, implying that this method may demand greater estimation accuracy. One can possibly solve it by increasing the number of observations N_R, $N_{\boldsymbol{\Psi}}$ or adopting a more robust technique for estimating $\widehat{\boldsymbol{\Psi}}_{tl}$ and $\widehat{\mathbf{R}}_{kl}$. In the next section, we will consider only the perfect knowledge of channel statistics to evaluate the scalable CF scheme in its full performance. Regarding the fine-tuning schemes, they were not also greatly affected by the imperfect knowledge of channel statistics.

6.1.5 Average Spectral Efficiency

From now on, we compare the performance of our proposed AP selection method with the only one that is also a scalable solution, the scalable CF scheme [26]. To compute the average SE, we consider only the LP-MMSE and P-MMSE since they provide the best interference mitigation for the distributed and centralized implementations, respectively. Besides, we are considering the perfect knowledge of channel statistics.

Fig. 17 shows the average achievable SE per UE as a function of the maximum number of UEs that each AP can serve (U_{max}) by varying U_{max} from 1 to 10. Higher values for U_{max} are not considered since it is limited to $\tau_p = 10$. As can be observed in Fig. 17, the proposed method can improve the average SE up to 96.4% for the distributed implementation when $U_{max} = 1$ and achieves the highest average SE when $U_{max} = 2$. One can note that for small values of U_{max} (such as $U_{max} = 2$) the additional amount of interference generated by the final AP cluster is still easily mitigated by the LP-MMSE. However, for $U_{max} > 6$, the interference levels increase even more, and the scalable CF has slightly better results. Despite this, Fig. 17 shows that the proposed method outperforms the scalable CF for all considered values of U_{max} in P-MMSE, improving up to 44.6% the average SE. Besides, Fig. 17 demonstrates that a proper U_{max} allows the network to achieve the best SEs for LP-MMSE and P-MMSE, in which case one should set $U_{max} = 2$ and $U_{max} = 8$, respectively. In Fig. 17, one can also note that the scalable CF scheme needs to employ APs with greater values of U_{max} to achieve similar results than ours with smaller U_{max}.

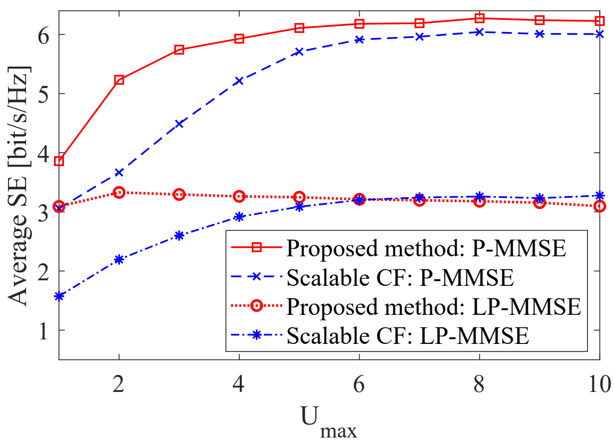


Figure 17 – Average DL SE versus U_{max}. Parameters setting: $L = 100$, $N = 1$, and $K = 20$. Perfect knowledge of channel statistics.

Fig. 18 shows the average SE versus the number of antennas per AP, N. Note that the SE of our solution overcomes the scalable CF scheme when both use $U_{max} = 4$ and match the scalable CF when $U_{max} = 10$. However, the proposed method slightly loses performance for $N \geq 4$, when $U_{max} = 10$ is employed. The explanation for the results in Fig. 18 is similar to those given for the previous ones. That is, a suitable U_{max} allows the proposed solution to generate less interference and increase SE, but an inappropriate U_{max} can degrade performance and make baseline solutions perform slightly better.

Fig. 19 presents the analysis regarding achievable average SE versus the number of UEs, K. For LP-MMSE, the proposed method improves the average SE for all values of K

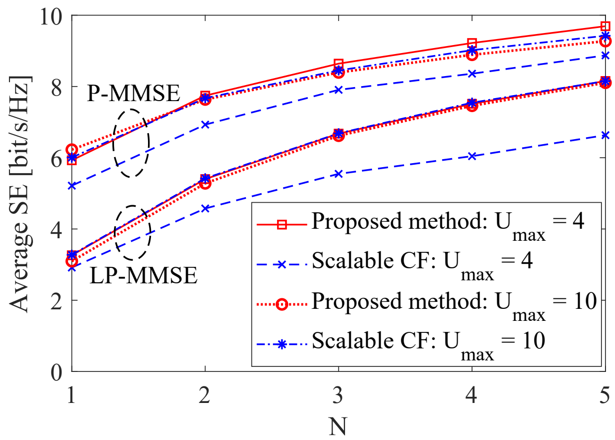


Figure 18 – Average SE as a function of the number of antennas per AP N. Parameters setting: $L = 100$ and $K = 20$. Perfect knowledge of channel statistics.

by at most 17% compared to the scalable CF scheme with $U_{max} = 4$. Besides, our scheme performs better with $U_{max} = 4$ than $U_{max} = 10$, being outperformed by the scalable CF scheme with $U_{max} = 10$ and $K = 80$, but the difference is negligible. For P-MMSE, the proposed method improves the average SE up to 47% when $K = 70$ and U_{max} equal to 4. Moreover, one can note that $U_{max} = 10$ is more suitable for our solution for $K > 25$, i.e., when the massive MIMO condition $(M/K > 4)$ is lost. This happens because the number of interfering UEs increases, while the number of APs in each AP cluster (L_k) decreases. Thus, for $K > 25$ and $U_{max} = 4$, our solution could not provide gains compared to the scalable CF method with $U_{max} = 10$. This is because in addition to the clusters decrease, more UEs are partially served by the same APs, worsening the P-MMSE performance.

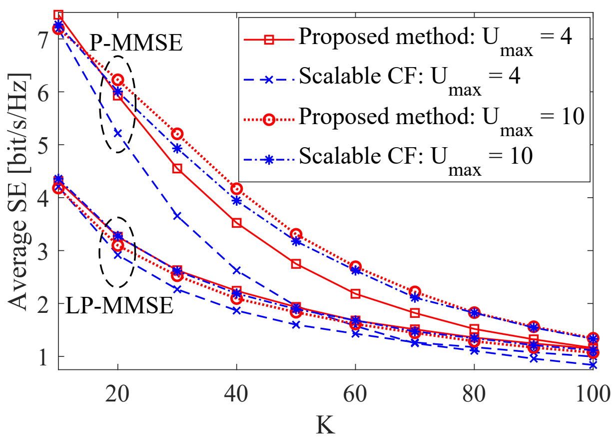


Figure 19 – Average SE versus the number of UEs K. Parameters setting: $L = 100$ and $N = 1$. Perfect knowledge of channel statistics.

Fig. 20 presents the average SE versus the number of APs, L, where U_{max} is equal to 4 or 10, and N decreases with L in order to keep $M = 100$. It is possible to observe that for small values of L the SEs are not too high due to the low macrodiversity, as the APs can serve only a few UEs in a coverage area of 1×1 km. However, as L increases and N decreases, the SE improves, with the behavior of the SEs for $U_{max} = 4$ and $U_{max} = 10$

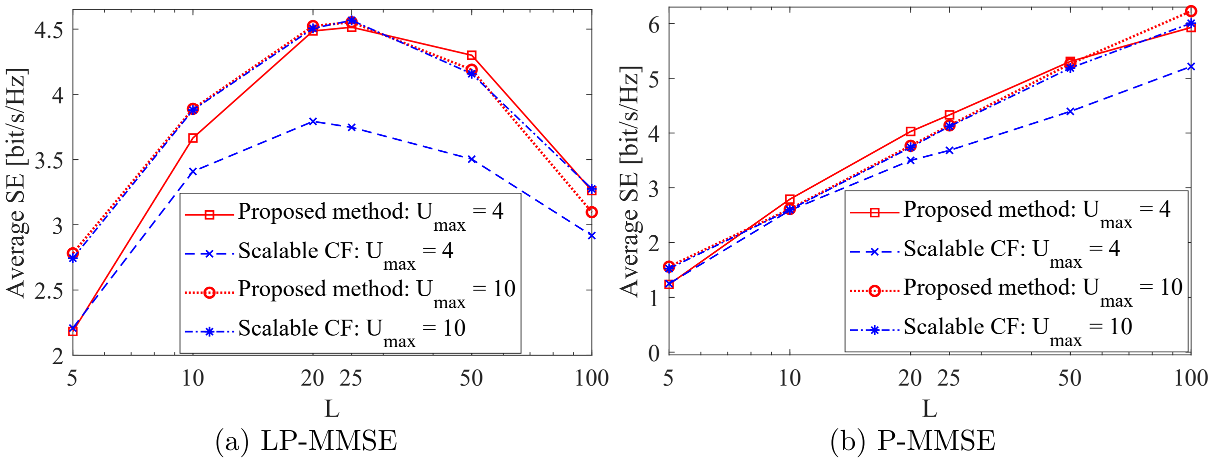


Figure 20 – Average SE versus the number of APs L. Parameters setting: $K = 20$ and N varies in order to keep M fixed at 100, where $M = NL$.

following the same explanation of previous results. In Fig. 20 (a), one can observe that the SE achieves a maximum value between $L = 20$ and $L = 25$. In this interval, the LP-MMSE could find the best balance between the strength of the received signal and interfering ones. However, for $L > 25$, the AP clusters are made up of more APs, consequently increasing interference and reducing the effectiveness of the LP-MMSE precoding. Additionally, Figs. 19 and 20 demonstrate that even if a system utilizes higher processing capacity APs, one can adapt U_{max} according to the network conditions to improve SE and reduce computational cost in some scenarios if it is used jointly with our solution.

6.2 Simulation Results: Chapter 4

This section presents the numerical results of a UC system utilizing the proposed method to reduce the effects of inter-CPU coordination, discussed in Chapter 4. The same parameters and assumptions discussed previously are adopted in this section but with some specificities, which are: the total transmission powers per UE and AP are 100 mW, 200 mW, respectively, and we have set $U_{\max} = \tau_p$. The AP selection scheme that jointly performs pilot assignment and AP clustering, described in Subsection 3.2.1 is utilized, which we have named as scalable cell-free (SCF) in this section. Besides, when it is used with the proposed method, we call it by SCF + CPU lim. We also compare the results with [10], where we adapted [10] to the proposed scenario. Specifically, it is considered that the primary CPU computes the UE channel sum gain regarding all APs linked to the primary CPU by fronthaul. If more than $\Lambda\%$ of the channel sum gain of the UE comes from the master AP, the UE is considered a cell edge UE. It is assumed that $\Lambda = 90\%$, and that [10] uses the SCF scheme to associate UEs and APs. In this section, only the perfect knowledge of channel statistics is considered.

Figs. 21, 22 and 23 depict the impacts of reducing inter-CPU coordination with the proposed method in a UC system composed of $L = 100$ APs and a variable number of UEs K. In Fig. 21, the proposed method ensures that the average SE is kept under small degradation, even if it limits the number of inter-coordinated UEs that each CPU can serve. The most significant decrease is observed for the LP-MMSE scheme, which is about 2.8% for $K = 50$. The SE is kept under small degradation because decreasing the number of inter-coordinated UEs per CPU also reduces the number of UEs that some APs serve (K_l). Hence, even though the UEs connect to fewer APs, they can also suffer less interference in the DL direction, helping precoding techniques with more modest interference mitigation capabilities, such as MR and LP-MMSE. Besides, although not shown in the figures, the SE of the 95% likely UEs (i.e., those presenting the minimum SE in the network) has been analyzed, and the proposed method revealed to affect the SE of the 95% likely UEs negligibly.

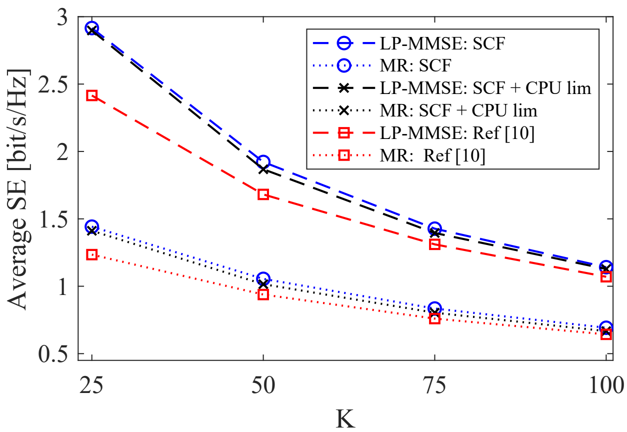


Figure 21 – Average DL SE achieved by varying the number of UEs K. Parameters setting: $L = 100$, $N = 1$, $J = 4$, and $K_{\text{int}} = \tau_p$, where $\tau_p = 10$.

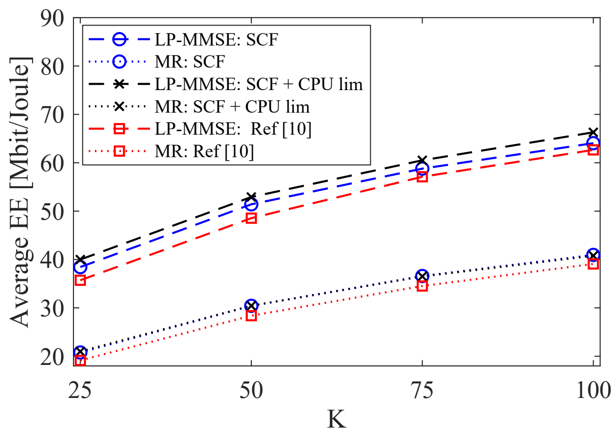


Figure 22 – Average DL EE achieved by varying the number of UEs K. Parameters setting: $L = 100$, $N = 1$, $J = 4$, and $K_{\text{int}} = \tau_p$, where $\tau_p = 10$.

In Fig. 22, the proposed method was able to slightly improve EE compared to the SCF scheme, reaching up to 4% improvement for $K = 25$. This is because decreasing K_l

allows the network to reduce the power consumption in each fronthaul link [24]. That is, the system achieves almost the same SE while consuming less power and increasing EE. On the other hand, the approach of [10] presented more SE losses than the proposed scheme and also losses in EE. This means that considering a hybrid approach between UC and a network-centric approach may lead the system to reduce its macro-diversity, resulting in performance losses.

In Fig. 23, observe that the average number of inter-coordinated UEs that each CPU serves grows with the number of UEs in a traditional UC system. On the other hand, it remains constant when using the proposed method, demonstrating that our solution allows UC systems to operate under a controlled inter-CPU coordination regime since it prevents the number of inter-coordinated UEs from growing with the number of UEs in each CPU. The approach proposed in [10] also reduces the number of inter-coordinated UEs in each CPU. However, it also grows with the number of UEs K, but slower than a traditional UC system.

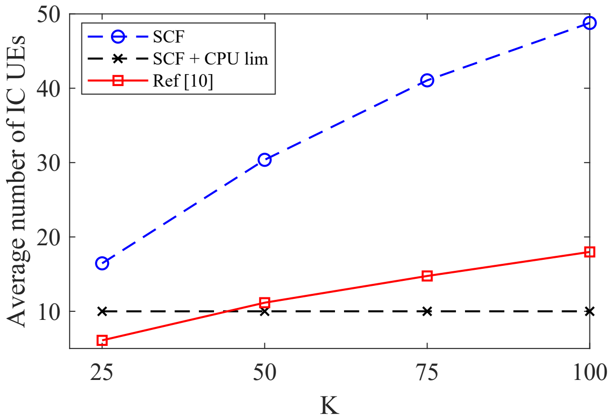


Figure 23 – Average number of inter-coordinated (IC) UEs per CPU. Parameters setting: $L = 100$, $N = 1$, $J = 4$, and $K_{\text{int}} = \tau_p$, where $\tau_p = 10$.

Additional results also reveal that the proposed method not only reduces K_l, but also decreases the number of APs connected to each UE (L_k). For instance, the fine-tuning strategy performed at the CPUs reduces the average L_k from 22.56 to 17.52, and the average K_l from 5.6 to 4.38, when $K = 25$. These quantities (K_l and L_k) are crucial in the CC of performing channel estimation and computing the precoding vectors [26]. They can also impact the traffic dependent power in 2.32, since it is proportional to K_l in the distributed implementation.

Fig. 24 shows that the proposed method also keeps the SE under minor degradation when the number of APs L varies. Furthermore, although not shown in the figures (to avoid redundancies), the EE was analyzed, and we can confirm that it is also improved when L varies. It is also possible to prevent the SE losses generated by the proposed method by employing fewer APs equipped with more antennas, as Fig. 25 illustrates. One can note that the proposed method does not impair the system's SE when $L \leq 25$. This is because the fewer APs in the coverage area, the further away the APs will be from the UE. Consequently, the AP cluster of the UE can comprise several APs with poor channel gains. Therefore, disconnecting some of these APs will not impact the UE's performance. A similar result is observed for $L \leq 60$ in Fig. 24. One can note that the approach proposed in [10] also presents performance losses in Figs. 24 and 25.

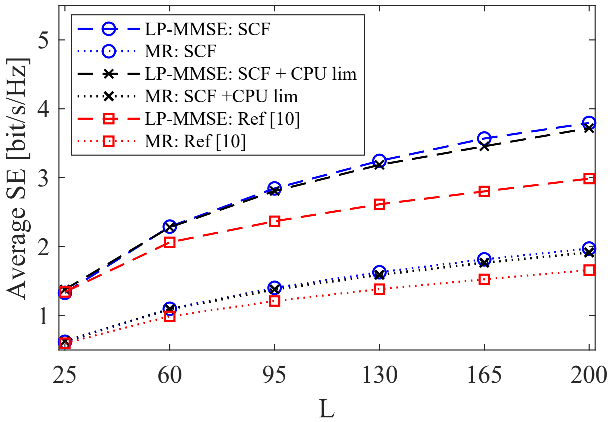


Figure 24 – Average DL SE achieved by varying L. Parameters setting: $K = 25$, $N = 1$, $J = 4$, and $K_{\text{int}} = \tau_p$.

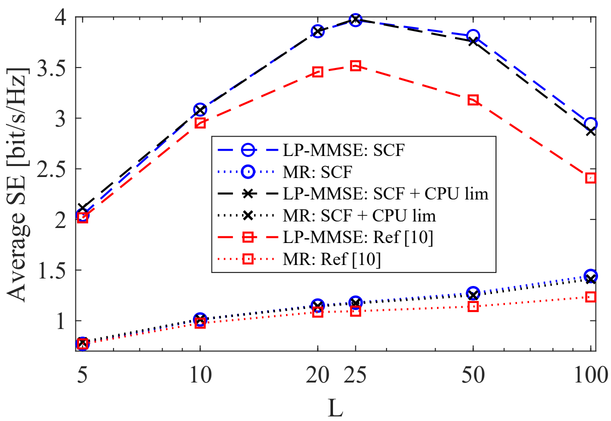


Figure 25 – Average DL SE achieved by varying L and N, while keeping $M = 100$. Parameters setting: $K = 25$, $J = 4$, and $K_{\text{int}} = \tau_p$.

Fig. 26 depicts the CDF of the DL SE by varying the number of UEs that each CPU can inter-coordinate, K_{int}. One can note that the variation of K_{int} does not affect the SE when $K_{\text{int}} > 10$, for $K = 25$. However, for $K_{\text{int}} \leq 10$, the system performance is compromised, specially by setting $K_{\text{int}} = 0$, which corresponds to a UC system without inter-CPU coordination. On the other hand, the variation of K_{int} has a negligible impact for a larger number of UEs, such as $K = 100$. This shows that the proposed method can maintain the network's performance under small degradation even if inter-CPU coordination is reduced significantly.

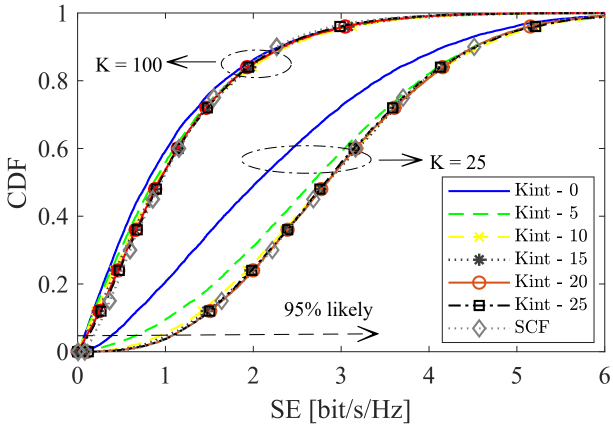


Figure 26 – Comparison of DL SE for different values of K_{int} using the LP-MMSE precoding scheme. Parameters setting: $L = 100$, $J = 4$, and $N = 1$.

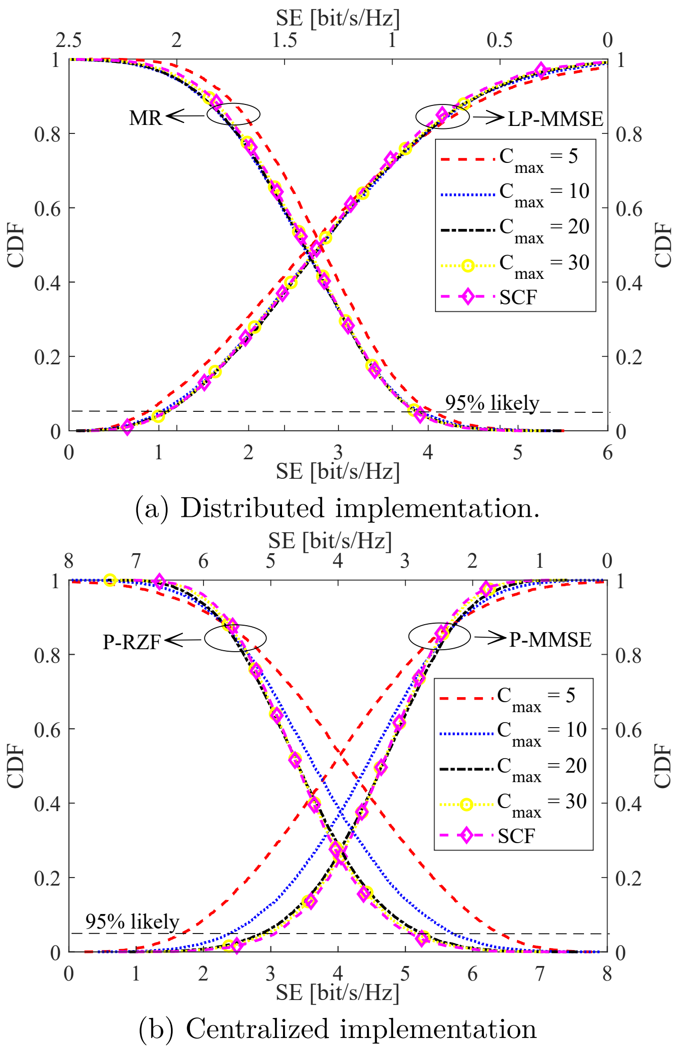
6.3 Simulation Results: Chapter 5

This section presents the numerical results associated with the proposed methods to limit UC systems' processing capacity and adjust AP clusters according to each network implementation, discussed in Chapter 5. This section also follows the same parameters and assumptions discussed previously but with some particularities, which are: the total transmission powers of the UEs and APs are 100 mW and 200 mW, respectively. We utilize a modified version of the SCF AP selection scheme in these simulations. Essentially, it is the same method, but the non-master APs do not require the UE to present a channel gain above a threshold value as in Section 3.2.1. Thus, the non-master APs serve the UE presenting the greatest channel gain in each pilot [27]. This approach makes the number of users the AP serves equal to the maximum value allowed, $K_l = U_{max}$. Therefore, it is possible to analyze how much the CC of the network can be reduced when using the proposed solutions since the CC is proportional to K_l. In this section, only the perfect knowledge of channel statistics is considered. The CC is computed in terms of the number of complex multiplications required to compute the precoding schemes for each UE in each coherence block, as depicted in Table 2.

6.3.1 Impacts of Limiting the Processing Capacity

We start by evaluating a network composed of $K = 25$ UEs and $L = 100$ APs equipped with $N = 1$ antenna. Fig. 27 presents the CDFs of the SE of UC systems with and without processing capacity limitation. It considers different processing capacity limitations, i.e., several values of C_{max}, and the system is compared with a traditional UC scheme (i.e., L_k is not restricted) that utilizes the AP selection strategy of [27], which we also have denoted as SCF. Besides, when the AP selection scheme of [27] is used with the proposed method, we call it by SCF + C_{max}.

In Fig. 27a, the SE is not as reduced by the variations of C$_{max}$. The SE even increases slightly for $10 \leq \mathrm{C}_{max} \leq 20$. This is because decreasing L_k also reduces K_l, helping precoding techniques such as LP-MMSE (of local processing) to mitigate interference. Still, this improvement has a limit since the SE decays about 9% when C$_{max}$ goes from 40 to 5. In Fig. 27b, the SE can suffer significant losses when C$_{max}$ is as small as 5. Hence, reducing the AP cluster sizes (L_k) may lead the centralized implementation to not exploit its full potential in mitigating interference and improving SE. Therefore, it is essential for this implementation to utilize more processing capacity, such as $\mathrm{C}_{max} \geq 20$.

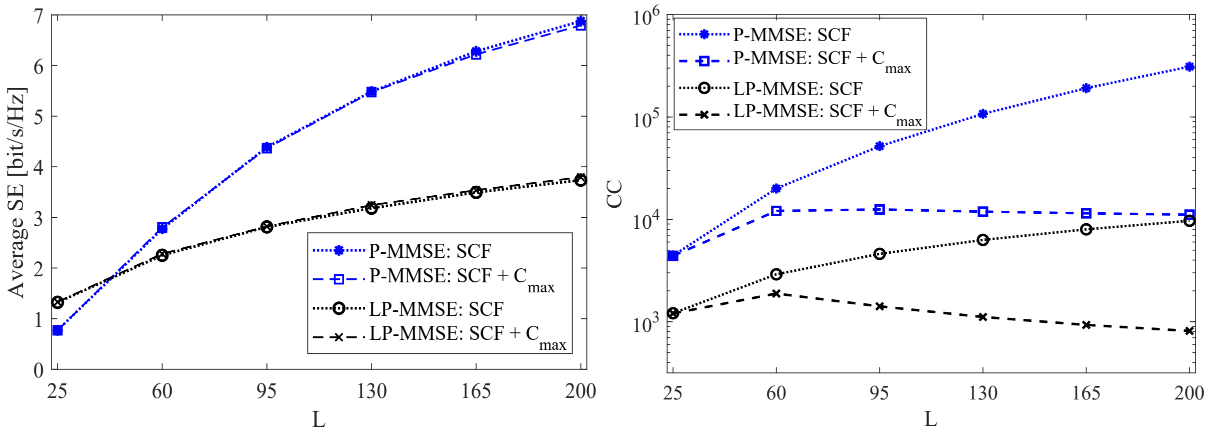


(a) Distributed implementation.

(b) Centralized implementation

Figure 27 – CDF of SE by varying C$_{max}$ from 5 to 30. Parameters setting: $L = 100$, $K = 25$, and $N = 1$.

Fig. 28 presents the SE and CC when the number of APs varies and by setting $K = 25$, and $\mathrm{C}_{max} = 20$. In Fig. 28a, the average SE grows with L for UC systems with and without processing capacity limitation. Despite this, limited systems have a significant advantage, as their CC does not always increase with L, starting to decay from $L = 60$. This behavior occurs because K_l reduces as L increases. Therefore, even if L_k remains constant, there will be a reduction in K_l, as Table 6 demonstrates. Additionally, it is possible to observe that the CC decreases by about $96,4\%$ when the processing capacity limitation is employed together with the P-MMSE for $L = 200$. However, a centralized implementation may require more processing capacity to be feasible compared to the distributed implementation. For instance, the P-MMSE scheme has a CC similar to LP-MMSE (without processing limitation) even limiting the processing capacity, when L is as large as 200.



(a) SE (b) CC

Figure 28 – Average DL SE (a) and CC (b) achieved by varying the number of APs L. Parameters setting: $K = 25$, $N = 1$, and $\mathrm{C}_{max} = 20$.

Fig. 29 presents the EE achieved in the distributed implementation considering different values of C_{max} and a UC system without processing capacity limitation. Note that the processing capacity limitation can provide a considerable improvement in the EE, especially for small values of C_{max}. For instance, the processing capacity limitation guarantees an increase of about 10% in EE for $\mathrm{C}_{max} < 30$. Besides, the EE grows by about 61% in the LP-MMSE and 36% in the MR, when C_{max} decreases from 40 to 5. This happens because reducing K_l also decreases the power consumption in each fronthaul link. Thus, even though the system presents SE losses when $\mathrm{C}_{max} = 5$, the reduction of power consumption in each fronthaul link compensates them, increasing the EE.

Table 6 – Average number of APs per UE (L_k) and UEs per AP (K_l) without and with AP cluster control. Parameters setting: $K = 25$, $N = 1$, and $\mathrm{C}_{max} = 20$.

Method	$L = 95$		$L = 200$	
	K_l	L_k	K_l	L_k
SCF	10	38	10	80
With C_{max}	5.25	19.98	2.5	20

6.3.2 Impacts of AP Cluster Adjustment

From now on, we will investigate the impacts of adjusting the AP clusters in UC systems. We will focus on UC systems without processing capacity limitation to assess the full benefits of the AP cluster adjustment in reducing CC. Furthermore, we will consider only the P-MMSE and LP-MMSE schemes as they provide the best interference mitigation in centralized and distributed implementations.

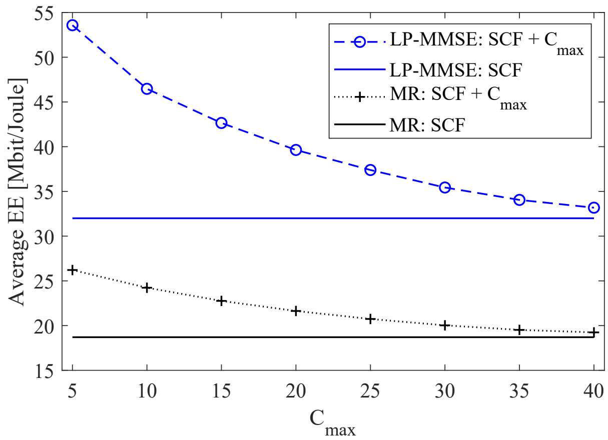
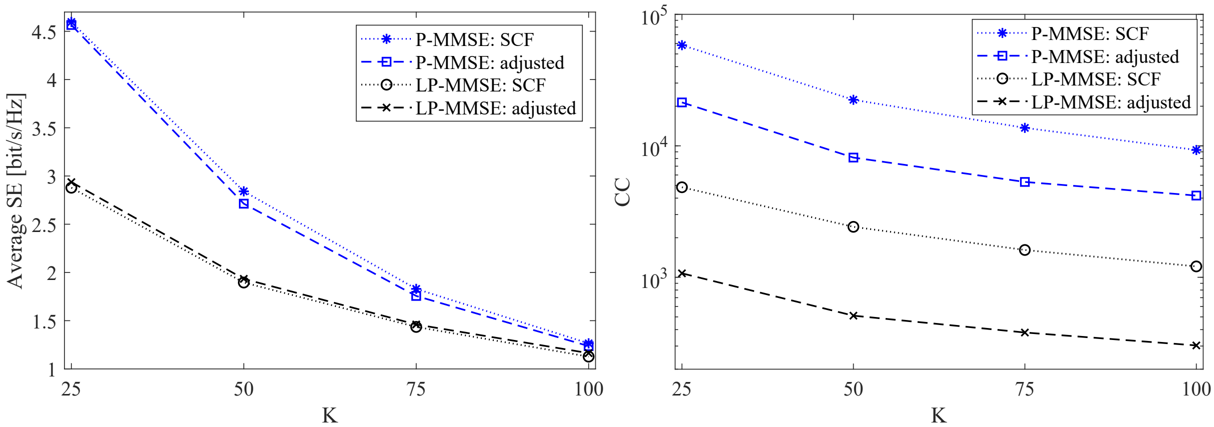


Figure 29 – Average EE achieved by varying C_{max}. Parameters setting: $L = 100$, $K = 25$, and $N = 1$.

Fig. 30 presents the average SE and CC versus the number of UEs K in a network composed of $L = 100$ APs equipped with $N = 1$ antenna. It can be noted that the proposed method causes a tiny reduction in the SE of P-MMSE. Despite this, the losses are not as expressive as in Fig. 27b. This is because the proposed method does not decrease L_k to a small value such as 5, as Table 7 indicates. One can also note that the proposed method causes a slight increase in the SE of LP-MMSE. Moreover, the AP cluster adjustment also reduces the CC of both network implementations, decreasing by up to 60% in the P-MMSE scheme for $K = 25$. Finally, the proposed method decreases K_l from 10 to 3.95 and L_k from 40 to 15.80, as illustrated in Table 7, indicating that the proposed strategy can also increase the EE in distributed implementation.



(a) SE (b) CC

Figure 30 – Average DL SE (a) and CC (b) achieved by varying the number of UEs K, when the proposed AP cluster adjustment is employed. Parameters setting: $L = 100$, $N = 1$.

Table 7 – Average number of APs per UE (L_k) and UEs per AP (K_l) without and with AP cluster adjustment. Parameters setting: $L = 100$ and $N = 1$.

Method	$K = 25$		$K = 50$	
	K_l	L_k	K_l	L_k
SCF	10	40	10	20
Distributed adjustment	4.32	17.3	4.38	8.75
Centralized adjustment	6.23	24.92	6.17	12.35

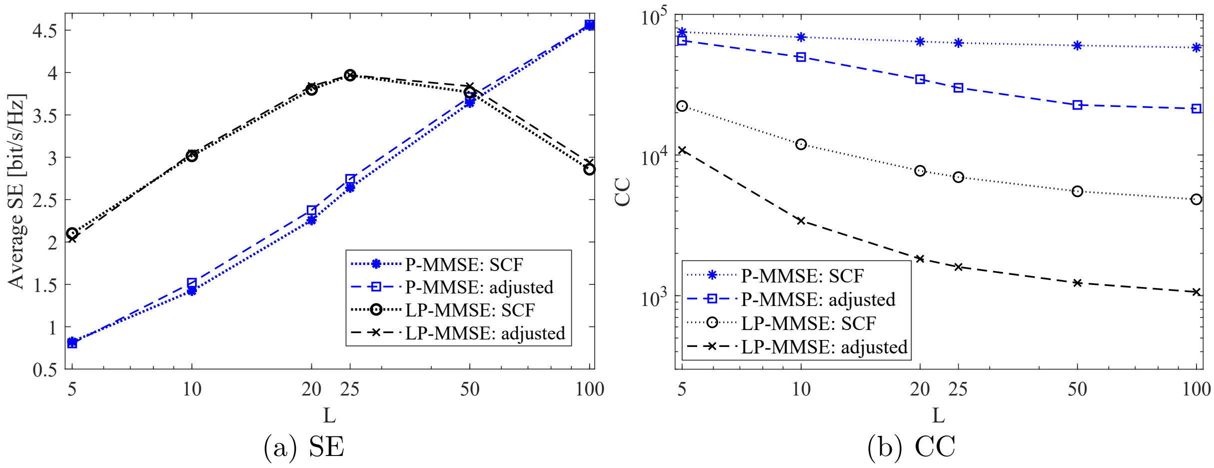


Figure 31 – Average DL SE (a) and CC (b) achieved by varying L and N, while keeping $M = 100$, when the proposed AP cluster adjustment is employed. Parameters setting: $K = 25$.

Fig. 31 presents the average SE and CC versus the number of UEs L and N for a fixed total number of antennas $M = LN = 100$ and setting the number of UEs to be $K = 25$. One can note that the same discussions about decreasing CC apply to this case. The difference is the SE behavior. When $L = 25$ and $N = 4$, the LP-MMSE scheme achieved the best balance regarding the amount of interference and desired signal, leading the average SE to its maximum value. Meanwhile, the P-MMSE presents better SE when the AP clusters are adjusted for $L < 100$. This is because the fewer APs in the coverage area, the further away the APs will be from the UE. Hence, the AP clusters can have many APs presenting poor channel gains. Therefore, disconnecting some of these APs will not impact the UE's performance. Additionally, it can be noticed that CC reduces as L increases and N decreases. The reduction is stronger in systems with AP cluster adjustment. At $L = 100$, the proposed method reduces the CC by about 63% and 78% for the P-MMSE and LP-MMSE schemes, respectively. Therefore the AP cluster adjustment can strongly reduce CC, especially for a large number of APs.

7 Conclusions and Future Works

This thesis presented solutions for enhancing the performance of UC CF massive MIMO systems, focusing on AP selection techniques. The first solution was a new scalable AP selection framework, which is an algorithm that exploits a matched-decision among the UEs and APs while guaranteeing connection for all UEs. The method consists of two stages, where the UEs first connect to an intermediate subset of APs and then form a final cluster. These steps aim to make the UEs and APs establish the best connection for both and then make the UEs expand their AP clusters intending to improve their SE. The method can improve the system's performance and afford scalability for baseline AP selection strategies. Three fine-tuning algorithms to be applied after the AP selection were also proposed. The first two fine-tuning algorithms are based on allocated power and SE. Results indicate that they can enable UEs to achieve almost the same SE as before while reducing the number of APs serving each UE. The third one aims to improve the total EE, and the results indicate that it can enhance the EE up to 43% for the LP-MMSE precoding method. Besides, AP selection schemes and fine-tuning algorithms were evaluated under perfect and imperfect knowledge of channel statistics. It is worth mentioning that all results achieved in this thesis were obtained under the assumptions and models presented in the document. Therefore, they present the network performance under specific conditions such as Rician fading and and error-free fronthaul.

The achievable SE in centralized and distributed network implementations were analyzed by varying the numbers of UEs K, APs L and antennas N per AP. Each AP could serve up to U_{max} UEs, and the APs could deal with different values for U_{max}. The results indicate that the proposed method (i.e., the matched-decision) can outperform baseline solutions and improve the SE of the worst UEs. For instance, the matched-decision can increase the SEs of the 95% likely UEs up to 163% and 100% in distributed and centralized implementations, respectively. The results also indicated that the scalable AP selection baseline method requires that the APs serve more UEs than our solution to enable the network to achieve similar SEs. Additional results revealed that U_{max} has to be set appropriately to the network condition, e.g., number of UEs and network implementation. Therefore, although an AP can serve more UEs, it can reduce U_{max} to improve the SE while reducing computational costs.

The second improvement is a novel method for reducing inter-CPU coordination in UC systems. The method considered that the number of inter-coordinated UEs that each CPU can serve is limited. Furthermore, it was assumed that the CPUs can drop inter-coordinated UEs presenting the weakest channel gains to reduce inter-CPU coordination. The proposed method was compared with two baseline schemes: a UC system without inter-CPU coordination reduction and a baseline scheme presenting it. The comparisons were carried out in terms of SE, EE, and the average number of inter-coordinated UEs that each CPU can serve. The results demonstrated that the proposed method allows the network to keep the SE under minor degradation, of at most 2.8%, even reducing inter-CPU coordination, compared to a traditional UC system. Our findings also indicated that the proposed method can provide slight improvements in EE while avoiding the number of inter-coordinated UEs per CPU becoming a function of the number of UEs. The same conclusions remain valid for networks with more APs and CPUs. We also observed that a small number of inter-coordinated UEs per CPU allows the system to provide SEs as high as a UC system without inter-CPU coordination control. Besides, the proposed method outperformed the baseline scheme that reduces inter-CPU coordination.

This thesis also investigated the performance of scalable UC CF massive MIMO systems whose processing capacity requirements do not increase with the number of APs. We analyzed UC systems whose AP clusters can have only a finite number of APs serving each UE. Furthermore, a method that adjusts the AP clusters to the network implementation was proposed. The results demonstrated that restricting the network processing capacity can improve the EE by up to 61%. However, it can degrade the SE of centralized implementation when the maximum number of APs serving the UE is small. On the other hand, AP clusters comprising just a few APs almost do not harm the SE of the distributed implementation. Simulation results also reveal that the proposed AP cluster adjustment can slightly improve the SE of distributed implementation while reducing the CC in both network implementations. For instance, the CC can decrease by up to 96% in centralized implementation. Finally, it is noteworthy that the results presented in this thesis are novel, useful for researchers working with AP selection methods, and can inspire new thesis and research papers on the theme.

7.1 Future Works

Some future works that can be derived from this thesis are described below:

- Expand our analyses to consider aspects such as limited fronthaul/backhaul capacity, non-reciprocity and hardware impairments. Besides, it is worthwhile to optimize the number of UEs that each AP can serve, and power allocation.

- Extending the analyzes regarding the effects of reducing inter-CPU coordination in centralized implementations, since the analyzes carried out in this thesis considered the impacts of inter-CPU coordination only in distributed implementation.

- Generate a model that quantifies the impacts of signaling demands on backhaul links due to inter-CPU coordination. This thesis presented the impacts of inter-CPU coordination in UC CF massive MIMO systems, but only in terms of the average number of inter-coordinated UEs per CPU. Thus, one can calculate the backhaul traffic as a function of CSI and data sharing.

- Calculate the CC by considering other network aspects such as reciprocity calibration, and discrete Fourier transform operations.
- Improve the EE modeling by considering other aspects of the network such as CC (in CPUs and APs) and backhaul traffic.
- Investigate the impacts of UE mobility and latency in UC CF massive MIMO. The analysis carried out in this thesis did not consider the effects of channel aging and the time required to exchange signals between CPUs and APs.

Publications

Patent and papers related to the thesis

- **Marx M. M. Freitas**, André Mendes Cavalcante, Daynara D. Souza, Luca Valcarenghi, João C. W. A Costa, Roberto M. Rodrigues, Gilvan S. Borges, Maria V. Marquezini, Igor Almeida, *"Method for Reduced Inter-CPU Coordination in User-Centric Cell-Free Massive MIMO Systems,"* Patent PCT/IB2022/060045, Ericsson Ltda., 2022.
- **Marx Freitas**, Daynara Souza, Gilvan Borges, André Mendes Cavalcante, Daniel Benevides da Costa, Maria Marquezini, Igor Almeida, Roberto Rodrigues, and João C. W. A. Costa, *"Matched-Decision AP Selection for User-Centric Cell-Free Massive MIMO Networks,"* in IEEE Transactions on Vehicular Technology, vol. 72, no. 5, pp. 6375-6391, May 2023.
- **Marx M. M. Freitas**, Daynara D. Souza, Daniel B. da Costa, André Mendes Cavalcante, Luca Valcarenghi, Gilvan S. Borges, Roberto Rodrigues, and João C. W. A. Costa, *"Reducing Inter-CPU Coordination in User-Centric Distributed Massive MIMO Networks,"* in IEEE Wireless Communications Letters, vol. 12, no. 6, pp. 957-961, Jun. 2023.
- **Marx M. M. Freitas**, Daynara D. Souza, A. L. P. Fernandes, Daniel B. da Costa, André Mendes Cavalcante, Luca Valcarenghi, and João C. W. A. Costa, *"Scalable User-Centric Distributed Massive MIMO Systems with Limited Processing Capacity,"* ICC 2023 - IEEE International Conference on Communications, Rome, Italy, Oct. 2023, pp. 4298-4304.

Papers submitted

- **Marx M. M. Freitas**, Daynara D. Souza, A. L. P. Fernandes, Daniel B. da Costa, André Mendes Cavalcante, Luca Valcarenghi, and João C. W. A. Costa, *"Scalable User-Centric Distributed Massive MIMO Systems with Restricted Processing Capacity,"* in IEEE Transactions on Wireless Communications.
- Daynara D. Souza, **Marx M. M. Freitas**, André L. P. Fernandes, Pedro H. J. Nardelli, Daniel Benevides da Costa, André Mendes Cavalcante, and João C. Weyl Albuquerque Costa, *"User-Centric Distributed Massive MIMO Systems Enabled by a Swarm of UAVs"*, in IEEE Transactions on Vehicular Technology.

- Michael A. S. Costa, **Marx M. M. Freitas**, Daynara D. Souza, André Mendes Cavalcante, Roberto M. Rodrigues, and João C. W. A. Costa, *"User-Centric Distributed Massive MIMO Systems: Is Scalability Beneficial for Indoor Environments?"* SBrT 2024 - XLII Brazilian Symposium on Telecommunications and Signal Processing, Belém, PA, Oct. 2024.

Remaining papers published during the period of the thesis

- Diogo Acatauassu, Moysés Licá, Aline Ohashi, André Lucas Pinho Fernandes, **Marx Freitas**, João C. W. A. Costa, Eduardo Medeiros, Igor Almeida, and André Mendes Cavalcante, *"An Efficient Fronthaul Scheme Based on Coaxial Cables for 5G Centralized Radio Access Networks,"* in IEEE Transactions on Communications, vol. 69, no. 2, pp. 1343-1357, Feb. 2021.
- **Marx Freitas**, Aline Ohashi, Diogo Acatauassu, João C. W. A. Costa, Eduardo Medeiros, Miguel Berg, Igor Almeida, and André Cavalcante, "Performance Evaluation of MGfast Systems Over Coaxial Cables,"in Journal of Communication and Information Systems, 36 (1), 52–61, 2021.
- Daynara D. Souza, **Marx M. M. Freitas**, Gilvan S. Borges, André Mendes Cavalcante, Daniel B. da Costa, and João C. W. A. Costa, *"Effective Channel Blind Estimation in Cell-Free Massive MIMO Networks,"* in IEEE Wireless Communications Letters, vol. 11, no. 3, pp. 468-472, Mar. 2022.
- Thaissa T. R. Ueoka, Daynara D. Souza, **Marx M. M. Freitas**, André Mendes Cavalcante, and João C. W. A. Costa, *"Performance of Cell-Free Massive MIMO Networks under Rayleigh and Rician fading,"* SBRT 2022 - XL Brazilian Symposium on Telecommunications and Signal Processing, Sta. Rita Do Sapucaí, Minas Gerais, Sept. 2022.
- Daynara D. Souza, **Marx M. M. Freitas**, Daniel B. da Costa, Gilvan S. Borges, André Mendes Cavalcante, Luca Valcarenghi, and João C. W. A. Costa, *"Effective Channel DL Pilot-Based Estimation in User-Centric Cell-Free Massive MIMO Networks,"* GLOBECOM 2022 - 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, Dec. 2022, pp. 705-710.
- Thaissa T. R. Ueoka, Michael A. S. da Costa, Daynara D. Souza, **Marx M. M. Freitas**, André L. P. Fernandes, Gilvan S. Borges, André Mendes Cavalcante, and João C. W. A. Costa, *"Pilot Allocation and Assignment Optimization in User-Centric Distributed Massive MIMO Networks."* SBrT 2023 - XLI Brazilian Symposium on Telecommunications and Signal Processing, São José Dos Campos, SP, Oct. 2023.

- Michael A. S. Costa, **Marx M. M. Freitas**, Thaissa T. R. Ueoka, Daynara D. Souza, André L. P. Fernandes, André Mendes Cavalcante, Roberto M. Rodrigues, and João C. W. A. Costa, *"A Fairer Comparison of Processing Implementations in User-Centric Distributed Massive MIMO Systems,"* SBrT 2023 - XLI Brazilian Symposium on Telecommunications and Signal Processing, São José Dos Campos, SP, Oct. 2023.

References

[1] 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, no. 6, pp. 1065–1082, Jun. 2014.

[2] Cisco, “Cisco Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update, 2017-2022 white paper,” 2019. [Online]. Available: https://www.ieee802.org/3/ad_hoc/bwa2/public/calls/19_0624/nowell_bwa_01_190624.pdf

[3] Ericsson, “Ericsson Mobility Report,” Nov. 2023. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report

[4] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, Nov. 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[5] M. Shirakabe, A. Morimoto, and N. Miki, “Performance evaluation of inter-cell interference coordination and cell range expansion in heterogeneous networks for lte-advanced downlink,” in *2011 8th International Symposium on Wireless Communication Systems*, Nov. 2011, pp. 844–848.

[6] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. Quek, and J. Zhang, “Enhanced intercell interference coordination challenges in heterogeneous networks,” *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 22–30, Jun. 2011.

[7] H.-S. Kim, D.-R. Kim, S.-H. Yang, Y.-H. Son, and S.-K. Han, “Mitigation of inter-cell interference utilizing carrier allocation in visible light communication system,” *IEEE Commun. Lett.*, vol. 16, no. 4, pp. 526–529, Apr. 2012.

[8] C. Kosta, B. Hunt, A. U. Quddus, and R. Tafazolli, “On interference avoidance through inter-cell interference coordination (icic) based on ofdma mobile systems,” *IEEE Commun. Surv. Tuts.*, vol. 15, no. 3, pp. 973–995, 2013.

[9] M. Simsek, M. Bennis, and İsmail Güvenç, “Learning based frequency- and time-domain inter-cell interference coordination in hetnets,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4589–4602, Nov. 2015.

[10] V. Ranjbar *et al.*, “Cell-free mMIMO support in the O-RAN architecture: A PHY layer perspective for 5G and beyond networks,” *IEEE Commun. Stand. Mag.*, vol. 6, no. 1, pp. 28–34, 2022.

[11] 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.

[12] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO.* Cambridge University Press, 2016.

[13] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, 2020.

[14] G. Interdonato, E. Björnson, H. Q. Ngo, P. K. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, p. 197, 2019.

[15] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983–2989, Jun. 2009.

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

[17] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No.01CH37202)*, vol. 3, May 2001, pp. 1745–1749 vol.3.

[18] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: a new architecture for future public wireless access," *IEEE Commun. Mag.*, vol. 41, no. 3, pp. 108–113, Mar. 2003.

[19] V. Sivarama, L. Angel, and V. Reinaldo, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, Apr. 2007, pp. 83–87.

[20] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G mobile and wireless communications technology.* Cambridge University Press, 2016.

[21] I. D. Garcia, N. Kusashima, K. Sakaguchi, and K. Araki, "Dynamic cooperation set clustering on base station cooperation cellular networks," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2010, pp. 2127–2132.

[22] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, “Cell-free massive MIMO systems,” in *2015 49th Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 695–699.

[23] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, “Cell-free massive MIMO versus small cells,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[24] 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. Green Commun. Netw*, vol. 2, no. 1, pp. 25–39, 2018.

[25] S. Buzzi and C. D’Andrea, “Cell-free massive MIMO: User-centric approach,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[26] E. Björnson and L. Sanguinetti, “Scalable cell-free massive MIMO systems,” *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

[27] Özlem Tugfe Demir, E. Björnson, and L. Sanguinetti, “Foundations of user-centric cell-free massive MIMO,” *Foundations and Trends® in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, Jan. 2021. [Online]. Available: http://dx.doi.org/10.1561/2000000109

[28] E. Björnson and L. Sanguinetti, “Making cell-free massive MIMO competitive with MMSE processing and centralized implementation,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2019.

[29] M. M. M. Freitas, D. D. Souza, D. B. da Costa, A. M. Cavalcante, L. Valcarenghi, and J. C. W. A. Costa, “Scalable user-centric distributed massive MIMO systems with limited processing capacity,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 4298–4304.

[30] M. M. M. Freitas, D. D. Souza, D. B. d. Costa, A. M. Cavalcante, L. Valcarenghi, G. S. Borges, R. Rodrigues, and J. C. W. A. Costa, “Reducing inter-CPU coordination in user-centric distributed massive MIMO networks,” *IEEE Wireless Commun. Lett.*, vol. 12, no. 6, pp. 957–961, 2023.

[31] M. A. S. da Costa, M. M. M. Freitas, T. T. R. Ueoka, D. D. Souza, A. L. P. Fernandes, A. M. Cavalcante, R. M. Rodrigues, and J. C. W. A. Costa, “A fairer comparison of processing implementations in user-centric distributed massive MIMO systems,” in *SBrT 2023 - XLI Brazilian Symposium onTelecommunications and Signal Processing*, Oct. 2023.

[32] T. T. R. Ueoka, M. A. S. da Costa, D. D. Souza, M. M. M. Freitas, A. L. P. Fernandes, G. S. Borges, A. M. Cavalcante, and J. C. W. A. Costa, "Pilot allocation and assignment optimization in user-centric distributed massive MIMO networks," in *SBrT 2023 - XLI Brazilian Symposium onTelecommunications and Signal Processing*, Oct. 2023.

[33] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, Jul. 2019.

[34] N. Rajatheva *et al.*, "White paper on broadband connectivity in 6G," Apr. 2020. [Online]. Available: https://arxiv.org/abs/2004.14247

[35] I. F. Akyildiz *et al.*, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, Jul. 2020.

[36] M. Alsabah, M. A. Naser, B. M. Mahmmod, S. H. Abdulhussain, M. R. Eissa, A. Al-Baidhani, N. K. Noordin, S. M. Sait, K. A. Al-Utaibi, and F. Hashim, "6G wireless communications networks: A comprehensive survey," *IEEE Access*, vol. 9, pp. 148 191–148 243, 2021.

[37] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[38] 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)*, 2019, pp. 554–558.

[39] S. Mosleh, H. Almosa, E. Perrins, and L. Liu, "Downlink resource allocation in cell-free massive MIMO systems," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, 2019, pp. 883–887.

[40] S. Chakraborty, E. Björnson, and L. Sanguinetti, "Centralized and distributed power allocation for max-min fairness in cell-free massive MIMO," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 576–580.

[41] 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.

[42] R. Nikbakht, R. Mosayebi, and A. Lozano, "Uplink fractional power control and downlink power allocation for cell-free networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 774–777, 2020.

[43] J. An and F. Zhao, "Trajectory optimization and power allocation algorithm in mbs-assisted cell-free massive MIMO systems," *IEEE Access*, vol. 9, pp. 30 417–30 425, 2021.

[44] Y. Zhang, H. Cao, P. Zhong, C. Qi, and L. Yang, "Location-based greedy pilot assignment for cell-free massive MIMO systems," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 392–396.

[45] D. D. Souza, M. M. M. Freitas, G. S. Borges, A. M. Cavalcante, D. B. da Costa, and J. C. W. A. Costa, "Effective channel blind estimation in cell-free massive MIMO networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 468–472, 2022.

[46] H. Q. Ngo and E. G. Larsson, "Blind estimation of effective downlink channel gains in massive MIMO," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2919–2923.

[47] ——, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, 2017.

[48] M. Sarker and A. O. Fapojuwo, "Granting massive access by adaptive pilot assignment scheme for scalable cell-free massive MIMO systems," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, 2021, pp. 1–5.

[49] S. Buzzi, C. D'Andrea, M. Fresia, Y.-P. Zhang, and S. Feng, "Pilot assignment in cell-free massive MIMO based on the hungarian algorithm," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 34–37, 2021.

[50] 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*, 2021, pp. 1–7.

[51] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive MIMO: A blessing in disguise," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5153–5169, 2019.

[52] D. D. Souza, M. M. M. Freitas, D. B. da Costa, G. S. Borges, A. M. Cavalcante, L. Valcarenghi, and J. C. Weyl Albuquerque Costa, "Effective channel DL pilot-based estimation in user-centric cell-free massive MIMO networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2022, pp. 705–710.

[53] J. Myung, K. Kim, and Y.-J. Ko, "Transmit power control game with a sum power constraint in cell-free massive MIMO downlink," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 632–635, 2022.

[54] K. Kim, J. Myung, Y.-J. Ko, and W. Y. Lee, “Beamforming and power optimization for user fairness in cell-free MIMO systems,” *IEEE Access*, vol. 11, pp. 51 032–51 046, 2023.

[55] J. Fang, P. Zhu, J. Li, F.-C. Zheng, and X. You, “Cell-free mMIMO systems in short packet transmission regime: Pilot and power allocation,” *IEEE Trans. Veh. Technol.*, pp. 1–16, 2024.

[56] Y. Gao, P. Cheng, Z. Chen, K. T. Phan, W. Xiang, and Y.-P. P. Chen, “A scalable energy consumption optimization in downlink cell-free massive MIMO,” *IEEE Wireless Commun. Lett.*, vol. 13, no. 2, pp. 397–401, 2024.

[57] M. Obeed, Y. Al-Eryani, and A. Chaaban, “Alternating channel estimation and prediction for cell-free mMIMO with channel aging: A deep learning based scheme,” in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 3590–3595.

[58] M. S. Oh, A. B. Das, S. Hosseinalipour, T. Kim, D. J. Love, and C. G. Brinton, “A decentralized pilot assignment algorithm for scalable o-ran cell-free massive MIMO,” *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 373–388, 2024.

[59] L. Zhang, S. Yang, and Z. Han, “Pilot assignment for cell-free massive MIMO: A spectral clustering approach,” *IEEE Wireless Commun. Lett.*, vol. 13, no. 1, pp. 243–247, 2024.

[60] S. Chen, J. Zhang, E. Björnson, J. Zhang, and B. Ai, “Structured massive access for scalable cell-free massive MIMO systems,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1086–1100, Apr. 2021.

[61] 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)*, Jul. 2019, pp. 1–6.

[62] H. T. Dao and S. Kim, “Effective channel gain-based access point selection in cell-free massive MIMO systems,” *IEEE Access*, vol. 8, pp. 108 127–108 132, Jun. 2020.

[63] T. H. Nguyen, T. K. Nguyen, H. D. Han, and V. D. Nguyen, “Optimal power control and load balancing for uplink cell-free multi-user massive MIMO,” *IEEE Access*, vol. 6, pp. 14 462–14 473, 2018.

[64] J. Shi, X. Chen, N. Huang, H. Jiang, Z. Yang, and M. Chen, “Power-efficient transmission for user-centric networks with limited fronthaul capacity and computation resource,” *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5649–5660, 2020.

[65] V. Ranasinghe, N. Rajatheva, and M. Latva-aho, “Graph neural network based access point selection for cell-free massive MIMO systems,” in *Proc. IEEE Global Commun. Conf.*, Feb. 2021, pp. 01–06.

[66] T. X. Vu, S. Chatzinotas, S. ShahbazPanahi, and B. Ottersten, “Joint power allocation and access point selection for cell-free massive MIMO,” in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[67] A. A. Polegre and A. G. Armada, “User-centric massive MIMO systems with hardening-based clusterization,” in *WSA 2020; 24th International ITG Workshop on Smart Antennas*, May 2020, pp. 1–5.

[68] V. Croisfelt, T. Abrão, and J. C. Marinello, “User-centric perspective in random access cell-free aided by spatial separability,” *IEEE Internet Things J.*, pp. 1–1, 2022.

[69] M. Freitas, D. Souza, G. Borges, A. M. Cavalcante, D. B. da Costa, M. Marquezini, I. Almeida, R. Rodrigues, and J. C. W. A. Costa, “Matched-decision AP selection for user-centric cell-free massive MIMO networks,” *IEEE Trans. Veh. Technol*, vol. 72, no. 5, pp. 6375–6391, 2023.

[70] X. Hu, C. Zhong, X. Chen, W. Xu, H. Lin, and Z. Zhang, “Cell-free massive MIMO systems with low resolution ADCs,” *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6844–6857, 2019.

[71] S. Shekhar, A. Subhash, M. Srinivasan, and S. Kalyani, “Joint power-control and antenna selection in user-centric cell-free systems with mixed resolution ADC,” *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8400–8415, 2022.

[72] G. Femenias and F. Riera-Palou, “Fronthaul-constrained cell-free massive MIMO with low resolution ADCs,” *IEEE Access*, vol. 8, pp. 116 195–116 215, 2020.

[73] X. Zhang, T. Liang, K. An, G. Zheng, and S. Chatzinotas, “Secure transmission in cell-free massive MIMO with RF impairments and low-resolution ADCs/DACs,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8937–8949, 2021.

[74] Y. Zhang, M. Zhou, Y. Cheng, L. Yang, and H. Zhu, “RF impairments and low-resolution ADCs for nonideal uplink cell-free massive MIMO systems,” *IEEE Syst. J.*, vol. 15, no. 2, pp. 2519–2530, 2021.

[75] Y. Zhang, L. Yang, and H. Zhu, “Cell-free massive mimo systems with low-resolution ADCs: The Rician fading case,” *IEEE Syst. J.*, vol. 16, no. 1, pp. 1471–1482, 2022.

[76] X. Ma, X. Lei, P. T. Mathiopoulos, K. Yu, and X. Tang, “Scalable cell-free massive MIMO systems with finite resolution ADCs/DACs over spatially correlated Rician fading channels,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 7699–7716, 2023.

[77] T. Marzetta and A. Ashikhmin, “MIMO system having a plurality of service antennas for data transmission and reception and method thereof,” 2011, US Patent. 8594215.

[78] J. Choi, D. J. Love, and P. Bidigare, “Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory,” *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 802–814, Mar. 2014.

[79] E. Björnson, E. G. Larsson, and T. L. Marzetta, “Massive MIMO: ten myths and one critical question,” *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[80] S. Kim and B. Shim, “FDD-based cell-free massive MIMO systems,” in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[81] A. Abdallah and M. M. Mansour, “Efficient angle-domain processing for FDD-based cell-free massive MIMO systems,” *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2188–2203, 2020.

[82] T. Han and D. Zhao, “On the performance of FDD cell-free massive MIMO with compressed sensing channel estimation,” in *2021 IEEE 21st International Conference on Communication Technology (ICCT)*, 2021, pp. 238–242.

[83] ——, “Downlink channel estimation in FDD cell-free massive MIMO,” *Physical Communication*, vol. 52, p. 101614, 2022.

[84] B. Hassan, S. Baig, and S. Aslam, “On scalability of FDD-based cell-free massive MIMO framework,” *Sensors*, vol. 23, no. 15, 2023.

[85] J. Morte Palacios, O. Raeesi, A. Gokceoglu, and M. Valkama, “Impact of channel non-reciprocity in cell-free massive MIMO,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 344–348, Mar. 2020.

[86] A. A. Ohashi, D. B. d. Costa, A. L. P. Fernandes, W. Monteiro, R. Failache, A. M. Cavalcante, and J. C. W. A. Costa, “Cell-free massive MIMO-NOMA systems with imperfect SIC and non-reciprocal channels,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1329–1333, Jun. 2021.

[87] O. Özdogan, E. Björnson, and E. G. Larsson, “Massive MIMO with spatially correlated Rician fading channels,” *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, May 2019.

[88] O. Özdogan, E. Björnson, and J. Zhang, "Performance of cell-free massive MIMO with Rician fading and phase shifts," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5299–5315, Aug. 2019.

[89] A. A. Polegre *et al.*, "Channel hardening in cell-free and user-centric massive MIMO networks with spatially correlated Ricean fading," *IEEE Access*, vol. 8, pp. 139 827–139 845, Jul. 2020.

[90] A. H. Jafari *et al.*, "Study on scheduling techniques for ultra dense small cell networks," in *Proc. IEEE VTC-Fall*, Sept. 2015, pp. 1–6.

[91] 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.

[92] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 974–978.

[93] 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.

[94] K. Upadhya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546–550, Apr. 2018.

[95] X. Wang, A. Ashikhmin, Z. Dong, and C. Zhai, "Two-stage channel estimation approach for cell-free IoT with massive random access," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1428–1440, 2022.

[96] F. Ye, J. Li, P. Zhu, D. Wang, H. Wu, and X. You, "Spectral efficiency analysis of cell-free distributed massive MIMO systems with imperfect covariance matrix," *IEEE Syst. J.*, pp. 1–11, 2022.

[97] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems is massive MIMO the answer?" *IEEE Trans. Wireless Commun*, vol. 14, no. 6, pp. 3059–3075, 2015.

[98] 3GPP, "5G; Study on channel model for frequencies from 0.5 to 100 GHz," *3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.901 Version 16.1.0 Release 16*, Nov. 2020.

[99] T. H. Nguyen, T. V. Chien, H. Q. Ngo, X. N. Tran, and E. Björnson, "Pilot assignment for joint uplink-downlink spectral efficiency enhancement in massive MIMO systems with spatial correlation," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8292–8297, 2021.

[100] T. Van Chien, E. Björnson, and E. G. Larsson, “Joint pilot design and uplink power allocation in multi-cell massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, 2018.

[101] M. Sarker and A. O. Fapojuwo, “Granting massive access by adaptive pilot assignment scheme for scalable cell-free massive MIMO systems,” in *2021 IEEE 93rd Veh. Technol. Conf. (VTC2021-Spring)*, 2021, pp. 1–5.