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

Daynara Dias Souza

Cell-Free Massive Multiple Antenna Systems: Estimation Techniques, Scalability Challenges, and Aerial Access Point Deployments

TD 17/2025

UFPA / ITEC / PPGEE
Guama University Campus
Belém – Pará – Brasil
2025

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

Daynara Dias Souza

Cell-Free Massive Multiple Antenna Systems: Estimation Techniques, Scalability Challenges, and Aerial Access Point Deployments

TD 17/2025

UFPA / ITEC / PPGEE
Guama University Campus
Belém – Pará – Brasil
2025

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)

D541c Dias Souza, Daynara.
Cell-Free Massive Multiple Antenna Systems : Estimation Techniques, Scalability Challenges, and Aerial Access Point Deployments / Daynara Dias Souza. — 2025.
72 f. : il. color.

Orientador(a): Prof. Dr. João Crisóstomo Weyl Abuquerque Costa
Coorientador(a): Prof. Dr. Pedro Henrique Juliano Nardelli
Tese (Doutorado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2025.

1. Cell-free massive MIMO. 2. Autonomous aerial vehicles. 3. Blind estimation. 4. Computational complexity. 5. Pilot allocation and assignment. I. Título.

CDD 384

Daynara Dias Souza

Cell-Free Massive Multiple Antenna Systems: Estimation Techniques, Scalability Challenges, and Aerial Access Point Deployments

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.

The thesis was written under an agreement for a dual-doctoral degree programme between Lappeenranta–Lahti University of Technology LUT, Finland and the Federal University of Pará (UFPA), Brazil and jointly supervised by supervisors from both universities.

Supervisor: Prof. Dr. João Crisóstomo Weyl Abuquerque Costa
Co-supervisor: Prof. Dr. Pedro H. Juliano Nardelli

UFPA / ITEC / PPGEE
Guama University Campus
Belém – Pará – Brasil
2025

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

"CELL-FREE MASSIVE MULTIPLE ANTENNA SYSTEMS: ESTIMATION TECHNIQUES, SCALABILITY CHALLENGES, AND AERIAL ACCESS POINT DEPLOYMENTS"

AUTORA: **DAYNARA DIAS SOUZA**

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 DOUTORA EM ENGENHARIA ELÉTRICA NA ÁREA DE TELECOMUNICAÇÕES.

APROVADO EM: 26/09/2025

BANCA EXAMINADORA:

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

Prof. Dr. Pedro H. Juliano Nardelli
(Coorientador – LUT/Finlândia)

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

Prof. Dr. Diego Lisboa Cardoso
(Avaliador Interno - PPGEE/ITEC/UFPA)

Prof. Dr. Roberto Menezes Rodrigues
(Avaliador Externo ao programa - FEEB/UFPA)

Dr. André Mendes Cavalcante
(Avaliador Externo – ERICSSON)

Prof. Dr. Rui Miguel Henriques Dias Morgado Dinis
(Avaliador Externo – UNL/Portugal)

VISTO:

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

Abstract

Cell-free (CF) massive multiple-input multiple-output (MIMO) networks are envisioned as promising technologies for next-generation wireless networks. These networks stand out for their high spectral efficiency, lower susceptibility to blocking and shadowing, and uniform performance among users. As the CF massive MIMO scenario spreads a large number of access points (APs) over a large area serving a much smaller number of user equipment (UE) devices, problems related to power control, signal precoding, resources, decoding strategies, and interference management will be key issues. Furthermore, to achieve a scalable and efficient CF massive MIMO system, a user-centric (UC) approach is used to define the subset of APs serving each UE device, such that the complexity and resource requirements of each AP must remain finite when the number of UE devices is infinite. In this regard, this doctoral dissertation aims to present scalable methodologies suitable for efficiently implementing UC CF massive MIMO networks. Specifically, radio resource allocation, pilot assignment methods, downlink (DL) pilot-based, and blind effective channel estimation schemes are developed. Furthermore, a dynamic strategy is proposed to choose the best channel estimation method for each UE device. The results show that the proposed solutions can improve the spectral efficiency of the 50%-likely UE devices by up to 91% compared with the solutions presented in the literature. This doctoral dissertation also investigates how to implement a massive UC CF network MIMO enabled by aerial APs, which are useful to improve wireless connectivity in ultradense environments or flexibly extend coverage in more isolated areas. For this purpose, a trajectory optimization of a swarm of cooperative unmanned aerial vehicles (UAVs) is proposed. The results show that it is possible to optimize UAV deployment and trajectory without knowledge of the UE locations, reaching up to a 47% increase in average spectral efficiency. Finally, this doctoral dissertation investigates UC CF massive MIMO systems with a limited processing capability. More specifically, AP cluster adjustment methods are proposed to guarantee that the computational complexity (CC) of signal processing does not increase with the number of APs. The results show that it is possible to reduce the CC by up to 96% while maintaining spectral efficiency with minimal degradation.

Keywords: Autonomous aerial vehicles, cell-free massive MIMO, blind estimation, computational complexity, DL pilot-based estimation, effective channel estimation, pilot allocation and assignment, spectral and energy efficiency.

Resumo

As redes *cell-free* (CF) *massive multiple-input multiple-output* (MIMO) são apontadas como uma das tecnologias promissoras para redes sem fio de próxima geração. Estas redes se destacam pela alta eficiência espectral, menor suscetibilidade a bloqueios e sombreamento, e desempenho uniforme entre os usuários. Como o cenário CF massive MIMO espalha um grande número de pontos de acesso (APs - *access points*) em uma grande área atendendo a um número muito menor de equipamentos dos usuários (UE - *user equipment*), problemas relacionadas ao controle de energia, pré-codificação de sinal, alocação de recursos, estratégias decodificação e gerenciamento de interferência serão questões-chave. Além disso, para obter um sistema CF massive MIMO escalável e eficiente, usa-se uma abordagem centrada nos usuários (UC - *user-centric*) para definir o subconjunto de APs que servem a cada UE, de modo que a complexidade e os requisitos de recursos para cada AP devem permanecer finitos quando o número de UE tende ao infinito. Nesse sentido, esta tese de doutorado objetiva apresentar metodologias escaláveis adequadas para alcançar uma implementação eficiente de redes UC CF massive MIMO. Especificamente, são desenvolvidos métodos de alocação de recursos de rádio e atribuição piloto, e esquemas de estimação de canal efetivo de DL cega e baseada em pilotos de DL. Também propõe-se uma estratégia dinâmica e adaptável às condições dos UE para escolher o melhor método de estimação de canal para cada um. Os resultados mostram que as soluções propostas podem melhorar a eficiência espectral dos 50%-*likely* UE em até 91% comparado com as soluções da literatura. Nesta tese de doutorado, também apresenta-se propostas de avanços relacionados à implantação de uma rede CF massive MIMO habilitado através de APs aéreos, que são úteis para melhorar a conectividade sem fio em ambientes ultra-densos ou para estender a cobertura em áreas mais isolados de uma maneira flexível. Para isto, é proposto uma otimização de trajetória de um enxame de veículos aéreos não tripulados (VANTs) cooperativos. Os resultados mostram que é possível otimizar a localização e trajetória dos VANTs sem a necessidade do conhecimento das posições do UE, alcançando um aumento de até 47% na média da eficiência espectral. Por fim, também foram investigados sistemas UC CF massive MIMO com capacidade de processamento limitada. Mais especificamente, são propostos métodos para ajustar o grupos de APs de forma que a complexidade computacional (CC) para processamento de sinais não aumente com o número de APs. Os resultados mostram que é possível reduzir a CC em 96% sem causar grandes degradações na eficiência espectral.

Palavras-chaves: Alocação e atribuição de piloto, *cell-free massive* MIMO, complexidade computacional, estimação de canal efetivo, estimação cega, estimação baseado em piloto de DL, eficiência espectral e energética, veículos aéreos autônomos.

Acknowledgments

This work was carried out within the Electrical Engineering Graduate Program, at the Federal University of Pará UFPA, Brazil, and at the Department of Electrical Engineering, LUT School of Energy Systems at Lappeenranta–Lahti University of Technology LUT, Finland, jointly supervised by supervisors from both universities as part of a dual-doctoral degree agreement.

This work was supported in part by the Innovation Center, Ericsson Telecomunicações Ltda., Brazil; in part by the National Council for Scientific and Technological Development (CNPq), and in part by Coordination for the Improvement of Higher Education Personnel (CAPES). This work was also supported in part by the Research Council of Finland via: (a) X-SDEN project n.349965 (b) EnergyNet Fellowship n.321265/n.328869 and (c) ECO-NEWS project n.358928, and partly by Jane and Aatos Erkko Foundation via STREAM project, EU MSCA "COALESCE" project n.101130739, and EU CEF project CAREWINGS-5G project n.101180459.

During the period of this research, I have been awarded a scholarship from the CAPES Foundation, an agency under the Ministry of Education of Brazil, to conduct part of my doctoral research as a visiting student at LUT from September 2022 to June 2023. I have also been awarded a grant from the LUT Research Foundation for international research visits in December 2023.

Firstly, I would like to express my deepest gratitude to my supervisor at UFPA, Professor João Weyl, who has always encouraged me to pursue an academic career, advised me to seek international research experiences, and guided me since my undergraduate studies. I also extend my sincere thanks to my supervisor at LUT, Professor Pedro Nardelli, for the invaluable support and mentorship he provided throughout my doctoral studies. It has been a privilege to be advised by both of them.

I am also profoundly grateful to Professor Daniel Benevides da Costa, whose insightful feedback has significantly contributed to the development of this work. I would also like to express my sincere gratitude to Professor Gilvan Borges and Dr. André Cavalcante for their valuable ideas and pertinent comments, which have helped shape the direction of this research. To my dear friends and research partners, Dr. Marx Freitas and Dr. André Fernandes, I extend my heartfelt thanks for the many fruitful discussions that evolved into meaningful collaborations.

Special thanks to the reviewers of this dissertation, Professor Diana Moya Osorio and Professor Rui Dinis, for taking the time to read this dissertation and for their constructive comments. I would like to thank Dr. Hanna Niemelä for her thorough proofreading of this doctoral dissertation and some of my publications.

My deepest gratitude to my mother, Francy Santos, my sister (in memorial), Deysi Souza, my husband, Claudomiro Sales, and my brother, Deyverson Souza, for their love and encouragement throughout my life.

Finally, I extend my heartfelt thanks and gratitude to my colleagues in the Applied Electromagnetism Laboratory at UFPA and the Cyber-Physical Systems Group at LUT. I am especially grateful to Arun Narayanan, Cássia Almeida, Daniel Gutierrez-Rojas, Dick Carrillo, Henry Ferreira, Jules Moualeu, Majid Hussain, Mehar Ullah, Mohamed Korium, Pedro Silva, Aline Ohashi, Roberto Rodrigues, Jasmine Araújo, and Waldeir Monteiro.

Daynara Dias Souza
August 2025
Lappeenranta, Finland

Contents

List of publications

This dissertation is based on the following papers. The rights have been granted by the publishers to include the papers in the dissertation. The same entries will also appear in the References in a standard format.

I. Souza, D. D., Freitas, M. M. M., Borges, G. S., Cavalcante, A. M., Costa, D. B. da, and Costa, J. C. W. A., Effective channel blind estimation in cell-free massive MIMO networks, *IEEE Wireless Commun. Lett.*, vol. 11, (3), pp. 468–472, Mar. 2022. DOI: 10.1109/LWC.2021.3132418

II. Souza, D. D., Freitas, M. M. M., Costa, D. B. da, Borges, G. S., Cavalcante, A. M., Valcarenghi, L., and Costa, J. C. W. A., 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. DOI: 10.1109/GLOBECOM48099.2022.10001150

III. Souza, D. D., Fernandes, A. L. P., Freitas, M. M. de, Costa, D. B. da, Cavalcante, A. M., Nardelli, P. H. J., and Costa, J. C. W. A., Effective channel hybrid estimation in User-Centric distributed massive MIMO networks, in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Mar. 2025, pp. 01–06. DOI: 10.1109/WCNC61545.2025.10978257

IV. Souza, D. D., Freitas, M. M. M., Fernandes, A. L. P., Nardelli, P. H. J., Costa, D. B. da, Cavalcante, A. M., and Costa, J. C. W. A., Trajectory optimization in user-centric distributed massive MIMO systems enabled by UAV swarms, *IEEE Transactions on Vehicular Technology*, vol. 74, (6), pp. 9252–9268, Jun. 2025. DOI: 10.1109/TVT.2025.3539120

V. Freitas, M. M. M., Souza, D. D., Fernandes, A. L. P., Costa, D. B. da, Mendes Cavalcante, A., Valcarenghi, L., and Weyl Albuquerque Costa, J. C., Scalable user-centric distributed massive MIMO systems with restricted processing capacity, *IEEE Trans. Wireless Commun.*, vol. 23, (12), pp. 19 933–19 949, Dec. 2024. DOI: 10.1109/TWC.2024.3491153

Author's contribution

Daynara D. Souza is the principal author and investigator in Publications I, II, III, IV. In Publication V, Dr. Marx M. M. Freitas was the principal author, and Daynara D. Souza contributed to the discussion of the initial methodology, proposal development, and writing and reviewing of the first manuscript draft and the final version.

Other relevant publications

- Freitas, M., Souza, D., Costa, D. B. da, Borges, G., Cavalcante, A. M., Marquezini, M., Almeida, I., Rodrigues, R., and Costa, J. C. W. A., Matched-decision AP selection for user-centric cell-free massive MIMO networks, *IEEE Trans. Veh. Technol.*, pp. 1–16, Jan. 2023. DOI: 10.1109/TVT.2023.3235980

- Freitas, M. M. M., Souza, D. D., Costa, D. B. da, Cavalcante, A. M., Valcarenghi, L., Borges, G. S., Rodrigues, R., and Costa, J. C. W. A., Reducing inter-CPU coordination in user-centric distributed massive MIMO networks, *IEEE Wireless Commun. Lett.*, pp. 1–5, 2023. DOI: 10.1109/LWC.2023.3252000

- Freitas, M. M. M., Souza, D. D., Fernandes, A. L. P., Costa, D. B. da, Cavalcante, A. M., Valcarenghi, L., and Costa, J. C. W. A., Scalable user-centric distributed massive MIMO systems with limited processing capacity, in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 1–7

- Fernandes, A. L. P., Souza, D. D., Costa, D. B. da, Cavalcante, A. M., and Costa, J. C. W. A., Cell-free massive MIMO with segmented fronthaul: Reliability and protection aspects, *IEEE Wireless Commun. Lett.*, pp. 1–1, 2022. DOI: 10.1109/LWC.2022.3166485

- Fernandes, A. L. P., Souza, D. D., Natalino, C., Tonini, F., Cavalcante, A. M., Monti, P., and Costa, J. C. W. A., A cost assessment methodology for user-centric distributed massive MIMO architectures, *IEEE Open j. Commun. Soc.*, vol. 5, pp. 3517–3543, 2024. DOI: 10.1109/OJCOMS.2024.3406374

- Ueoka, T. T. R., Souza, D. D., Freitas, M. M., Cavalcante, A. M., and Costa, J. W., Performance of cell-free massive MIMO networks under Rayleigh and Rician fading, in *Anais do XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2022. DOI: 10.14209/sbrt.2022.1570811228

- Ueoka, T. T. R., Costa, M. A. S., Souza, D. D., Freitas, M. M., Fernandes, A., Borges, G. S., Cavalcante, A. M., and Costa, J. W., Pilot allocation and assignment optimization in user-centric distributed massive MIMO networks, in *Anais do XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2023. DOI: 10.14209/sbrt.2023.1570923555

- Costa, M. A. S., Freitas, M. M., Ueoka, T. T. R., Souza, D. D., Fernandes, A., Cavalcante, A. M., Rodrigues, R. M., and Costa, J. W., A fairer comparison of processing implementations in user-centric distributed massive MIMO systems, in *Anais do XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2023. DOI: 10.14209/sbrt.2023.1570915442

Nomenclature

Abbreviations

5G fifth-generation
AP access point
B5G beyond 5G
BS base station
CF cell-free
CPU central processing unit
CSI channel state information
CDF cumulative distribution function
CC computational complexity
CoMP coordinated multipoint
DL downlink
DCC dynamic cooperation clustering
EE energy efficiency
InH-open indoor hotspot open office
LoS line-of-sight
LP-MMSE local partial MMSE
L-MMSE local MMSE
MIMO multiple-input multiple-output
MMSE minimum mean square error
MSE mean square error
MR maximum ratio
NLoS non-line-of-sight
NMSE normalized mean square error
P-MMSE partial MMSE
P-RZF partial regularized zero-forcing
PSO particle swarm optimization
QoS quality-of-service
QoE quality-of-experience
RU radio unit
SE spectral efficiency

SINR signal-to-interference-plus-noise ratio

SNR signal-to-noise ratio

TRP transmission and reception point

TDD time-division duplex

UatF use-and-then-forget

UAV unmanned aerial vehicle

UE user equipment

UL uplink

UC user-centric

ULA uniform linear array

Notation

$\boldsymbol{a}$	Boldface lowercase letters denote column vectors
$\boldsymbol{A}$	Boldface uppercase letters denote column matrices
$\boldsymbol{I}_N$	$N \times N$ identity matrix
$(.)^{\mathrm{H}}$	Conjugate-transpose operation
$(.)^*$	Conjugate operation
$\lvert\mathscr{A}\rvert$	Cardinality of the set $\mathscr{A}$ operation
$\mathrm{diag}(\boldsymbol{A}_n)$	Block-diagonal matrix with square matrices on the diagonal
$\mathrm{tr}(.)$	Trace of a matrix operator
$\lVert . \rVert$	Euclidean norm operator
$\mathbb{E}\{.\}$	Expectation operator
$\mathscr{N}_{\mathbb{C}}\left(\mu, \sigma^2\right)$	Complex Gaussian random variable with mean μ and variance σ^2
$\mathscr{U}[a,b]$	Uniform random variable in the interval $[a,b]$

Symbols

L	Number of APs
N	Number of antennas per AP
K	Number of UEs
M	Total number of antennas in the network
τ_c	Number of complex-valued samples in each coherence block
τ_{up}	Number of complex-valued samples for UL pilots
τ_{dp}	Number of complex-valued samples for DL pilots
τ_d	Number of complex-valued samples for DL data
τ_u	Number of complex-valued samples for UL data
λ	Transmission wavelength
v	UE velocity
T_d	Delay spread

$\boldsymbol{h}_{kl}$	Channel vector between the UE k and AP l
$\boldsymbol{R}_{kl}$	Covariance matrix
κ_{kl}	Rician factor
φ_{lk}	Azimuth angle between AP l and UE k
θ_{lk}	Elevation angle between AP l and UE k
d_{lk}	Distance between AP l and UE k
β_{kl}	Large-scale fading coefficient
ψ_{kl}	Random phase shifts in the LoS component
p_{LoS}	LoS propagation probability
$\mathscr{M}_k$	Subset of APs that serve UE k
$\mathscr{D}_l$	Subset containing the UE served by AP l
$\boldsymbol{D}_{kl}$	Diagonal matrix that determines if AP l serves UE k
η_k	UL pilot transmit power of UE k
t_k	Index of the UL pilot assigned to UE k
$\mathscr{P}_k$	Subset of UEs that use the same pilot as UE k
$\boldsymbol{\Psi}_{t_k l}$	Correlation matrix of UE devices sharing pilot t_k
$\widehat{\boldsymbol{h}}_{kl}$	UL channel estimate of $\boldsymbol{h}_{kl}$
$\boldsymbol{n}_{t_k l}$	Additive UL noise
σ^2_{ul}	UL noise power
q_k	DL data symbol sent to UE k
ρ_{kl}	DL transmit power that AP l assigns to UE k
$\boldsymbol{x}_l$	DL data signal sent by access point (AP) l to its serving UEs
$\boldsymbol{w}_{kl}$	DL precoding vector
$y_{d,k}$	DL received signal of user equipment (UE) k
$n_{d,k}$	Additive DL noise
σ^2_{dl}	DL noise power
α_{kk}	DL effective channel gain for UE k
$\alpha_{kk'}$	DL effective interfering channel gain between UE k and k'
$\boldsymbol{v}_{kl}$	UL combining vector
$\mathscr{S}_k$	Subset of UE devices partially served by the same APs as UE k
$\text{SE}_k^{\text{sCSI}}$	Achievable DL SE of UE k assuming statistical CSI knowledge
$\text{SINR}_k^{\text{sCSI}}$	DL SINR of UE k assuming statistical CSI knowledge
$\text{SE}_k^{\text{eCSI}}$	Achievable DL SE of UE k assuming a method for CSI estimation
$\text{SINR}_k^{\text{eCSI}}$	DL SINR of UE k assuming a method for CSI estimation
$\text{SE}_k^{\text{pCSI}}$	Achievable DL SE of UE k assuming perfect CSI knowledge
$\text{SINR}_k^{\text{pCSI}}$	DL SINR of UE k assuming perfect CSI knowledge
SE_k	Achievable DL SE of UE k, without specifying CSI acquisition method
EE_{tot}	Total energy efficiency
P_{tot}	Total power consumption
B	System radio bandwidth
CHD_k	Channel hardening degree of UE k
FVP_k	Favorable propagation degree of UE k
ξ_k	Sample average power of the received signal at UE k
z_k	Index of the DL pilot assigned to UE k

Declarations

Turnitin

The originality of this dissertation has been reviewed with the Turnitin similarity checking service.

AI use

During the preparation of this doctoral dissertation, Daynara D. Souza, the author of the dissertation, used Grammarly in order to develop the presentation and grammar of textual information in the summary of the dissertation. After using Grammarly, the author reviewed and edited the content and takes full responsibility for the content of the doctoral dissertation.

1 Introduction

Wireless communications networks have evolved significantly over the years to allow for different applications that have become essential to society. A crucial question is how to further develop current wireless communication technologies to meet the constantly growing demand, thus avoiding a data traffic crisis and meeting rising expectations for quality of service. Industrial and academic researchers must design revolutionary new wireless networking technologies to keep up with an exponentially growing traffic rate, provide ubiquitous connectivity, and also quickly address demand surges and service outages. Key strategies for increasing the data transmission capacity of wireless networks include network densification, increasing the bandwidth, deployment of a large number of transmitting antennas to achieve a massive multiple-input multiple-output (MIMO) regime, and the use of advanced signal processing techniques that can mitigate interference. Among the potential solutions for beyond 5G (B5G) networks, the cell-free (CF) massive MIMO combines the advantages of the aforementioned approaches and has received considerable attention from the research community. Moreover, unmanned aerial vehicle (UAV)-aided communication is a crucial solution to address periods of intense demand and service interruptions.

1.1 Motivation

User-centric (UC) CF massive MIMO networks are envisioned as one of the promising technologies for next-generation wireless networks. Unlike traditional systems, in which user equipment (UE) connects to a specific cell tower, it allows the UE to communicate with multiple antennas spread across different locations [9], [10]. From the perspective of UE devices, their performance is consistent regardless of their location in the coverage area, effectively addressing the severe cell-edge interference of traditional systems. This concept gives rise to the new paradigm of "cell-free" networks. Furthermore, these networks leverage massive MIMO technology, employing many antennas to serve fewer UE devices, thereby enabling high data rates and simultaneous service for multiple UE devices [11], [12].

As the UC CF massive MIMO scenario spreads many access points (APs) over a large area serving a much smaller number of UE devices, it is essential to define the subset of APs serving each UE device properly. For this purpose, a "user-centric" approach is employed to achieve a scalable and efficient UC CF massive MIMO system such that the complexity and resource requirements for each AP must remain manageable even as the number of UE devices tends toward infinite [13]–[15]. While the benefits of UC CF massive MIMO networks are significant, it is essential to design resource allocation algorithms that allow for scalable implementation and avoid high computational complexity, signaling overhead, and energy consumption. This includes approaches for power control, signal precoding, resource allocation, AP selection schemes, decoding strategies, and interference management.

Specifically, the chosen AP selection approach plays a significant role in allowing for data sharing and resource allocation tasks to be confined within a few groups of APs, consequently limiting the required computational complexity on the central processing unit (CPU) and signaling overhead on the fronthaul and backhaul links. Moreover, it is also essential to perform resource allocation and decoding schemes based only on information that does not vary rapidly, such as large-scale fading, in order to limit the amount and frequency of information exchanged in the fronthaul and backhaul links [16].

A critical aspect of UC CF massive MIMO systems is that the channel hardening may be less pronounced than in cell-based systems due to the geographical distribution of the APs as well as by the fact that the UE connects only to a subset of APs equipped with few antennas [17]. Essentially, the channel hardening makes random channels behave almost deterministically when the number of antennas is large, enabling downlink (DL) data decoding based on statistical channel state information (CSI) of the effective channel [13], [15]. The *effective channel* is a scaling factor containing the channel distortion on the data symbol received by a given UE device. The effective channel can be analytically defined as the summation of the inner product between its channel vector and its precoding vector from all its serving APs. The UE devices need to know their effective channels to decode the corresponding DL data symbols coherently. The low degree of hardening in UC CF massive MIMO systems may lead to the need for the UE to perform a more reliable method for effective channel estimation [17].

In specific scenarios, a terrestrial wireless communication infrastructure might not be deployed, and it may malfunction or be unable to serve the UE demand. In those situations, aerial communication systems based on UAVs are regarded as a promising new paradigm to facilitate fast and highly flexible deployment of communication infrastructure due to their high maneuverability [12], [18]. UAVs are reusable aircraft designed to operate without an onboard pilot or passengers, which can be either piloted remotely or programmed to fly autonomously [19]. A viable solution to boost the performance of a UAV-aided communication system is to allow for UAVs to work collaboratively to achieve a collective objective, forming a fleet called swarm. Specifically, by leveraging the methodology from UC CF massive MIMO networks, the cooperative efforts of multiple distributed UAVs can realize the benefits of utilizing multiple antennas. The potential implementation of a swarm of UAVs as APs for UC CF massive MIMO networks introduces several key research challenges, including deployment and trajectory optimization in dynamic scenarios, determining the required wireless fronthaul capacity for UAVs as APs, and modeling power consumption and energy efficiency, taking into account parameters specific to UAV communication systems.

1.2 Objectives and research questions

The main objectives of this doctoral dissertation are to develop and evaluate efficient resource allocation and management strategies in UC CF massive MIMO networks, fo-

cusing on two themes: (*i*) developing methods for channel estimation that are adaptable to changes in the propagation environment and scalable, as well as strategies for resource allocation and assignment; and (*ii*) developing deployment and trajectory strategies for a UC CF massive MIMO network formed by a collaborative swarm of UAVs as aerial APs. The objectives of this dissertation can be further detailed as follows:

- Evaluate the performance of different effective channel estimation methods to assist in decoding DL data in UC CF massive MIMO systems;
- Develop and evaluate new DL pilot allocation and assignment algorithms for UC CF massive MIMO systems;
- Develop and evaluate a hybrid effective channel estimation method where each UE device estimates its effective channel using the conventional method called statistical CSI, blind estimation, or DL pilot-based training;
- Evaluate scalable CF massive MIMO systems using aerial APs, considering appropriate channel models, user mobility, and diverse scenarios for test cases;
- Develop and evaluate UAV trajectory optimization to maximize its energy efficiency in a scenario where the energy-limited UAVs act as aerial APs in CF massive MIMO networks;
- Develop scalable AP selection strategies that constrain computational complexity and signaling overhead for UC CF massive MIMO systems.

To achieve these objectives, this doctoral dissertation aims to answer the following research questions:

RQ 1: *Is it detrimental to rely on statistical CSI for DL data decoding in UC CF massive MIMO systems?*

The impact of relying on the channel hardening property and using statistical CSI in UC CF massive MIMO is addressed in RQ 1. In this respect, a blind estimation algorithm to be performed by the UE for DL data decoding can be proposed as an alternative solution to the conventional approach.

RQ 2: *What type of effective channel estimation should be used for DL data decoding in UC CF massive MIMO systems?*

RQ 2 focuses on investigating different channel estimation algorithms for DL data decoding in UC CF massive MIMO systems. In this way, DL pilot-based channel estimation can be compared with blind and statistical CSI approaches. To ensure a fair comparison,

DL pilot assignment has to be performed appropriately.

RQ 3: *How can a dynamic channel estimation approach be developed for UC CF massive MIMO networks?*

The choice between using statistical CSI, blind estimation, and DL pilot estimation may not be straightforward. RQ 3 assumes that it is possible to design a dynamic channel estimation method that adapts to the needs of UE and that can improve data rate performance in UC CF massive MIMO networks.

RQ 4: *How can the positioning and trajectories of UAVs be optimized with a relatively low complexity in UC CF massive MIMO systems?*

RQ 4 assumes that joint transmission based on UC CF massive MIMO systems can provide significant gains in terms of system capacity and coverage compared with traditional solutions. UAVs can synchronize their transmission using a wireless communication link to CPUs and with each other. By using appropriate radio resource management strategies, the required amount of overhead for signaling and the payload data can be maintained below the capacity of the fronthaul (i.e., the communication link among UAVs and CPUs). It is possible to design optimization algorithms for the trajectory or positioning of UAVs according to the objective.

RQ 5: *How can it be ensured that the computational complexity, signaling overhead, and energy consumption are kept under control in UC CF massive MIMO systems?*

RQ 5 investigates AP selection strategies for UC CF massive MIMO systems composed of multiple CPUs. RQ 5 assumes that it is possible to prevent the computational complexity (CC) from growing with the number of APs by controlling the maximum AP cluster size of each UE device.

1.3 Contributions and overview of chapters

Chapter 2 presents fundamental concepts of cellular and CF massive MIMO networks. It introduces the theoretical basis of this work, including the CF system model, channel modeling, channel estimation, AP selection, precoding techniques, power allocation, and DL data transmission processes. Chapter 3 answers the research questions of this doctoral dissertation. Chapter 4 delivers the final remarks and suggests directions for future work.

The contributions of this doctoral dissertation are presented in Chapter 3. In Section 3.1, a blind algorithm for effective channel estimation for UC CF massive MIMO networks is evaluated. Section 3.2 addresses scalable pilot allocation and assignment methods for DL pilot-based estimation that minimize pilot contamination effects. In Section 3.3, a method is proposed to dynamically adapt the selected estimation method for each UE device; i.e.,

a hybrid method that can use blind, DL pilot-based training, and statistical CSI methods. Section 3.3.5 addresses the trajectory optimization problem of UAVs jointly cooperating as the APs of a CF network under user mobility scenarios. In Section 3.5, a scalable AP selection method is proposed for adjusting the UE clustering aiming to restrict CC for signal processing as the number of APs grows.

2 User-centric cell-free massive MIMO networks

This chapter presents the theoretical foundation of this doctoral dissertation related to UC CF massive MIMO networks. It defines the key concepts of the cellular massive MIMO network and describes the motivation behind the CF concept. The system model for the UC CF massive MIMO system is presented, including channel models, estimation methods, and precoding schemes.

2.1 Cellular massive MIMO networks

Wireless communication systems have evolved over the years by exploiting the concept of cellular networks. In this topology, the coverage area is divided into nonoverlapping cells that operate independently using fixed-location base stations (BSs). Dividing the coverage area into cells is needed to deliver services with sufficiently high received signal energy over wide coverage areas, because the signal energy spreads out and less energy reaches a desired receiver as the distance increases [15], [20].

As illustrated in Figure 2.1, current wireless cellular networks are heterogeneous, consisting of overlapping cells, and they can be classified into two main tiers: coverage tier and hotspot tier. The coverage tier accounts for the outdoor BSs, provides coverage for a wide area, supports UE mobility, and serves many UE devices. The hotspot tier may consist of indoor or outdoor BSs, and it offers high throughput in small local areas to a few UE devices. There may be a coverage overlap between cells in different tiers, but not between cells in the same tier. In both cases, each UE device is connected to only one of the BSs to receive its DL and uplink (UL) services [20].

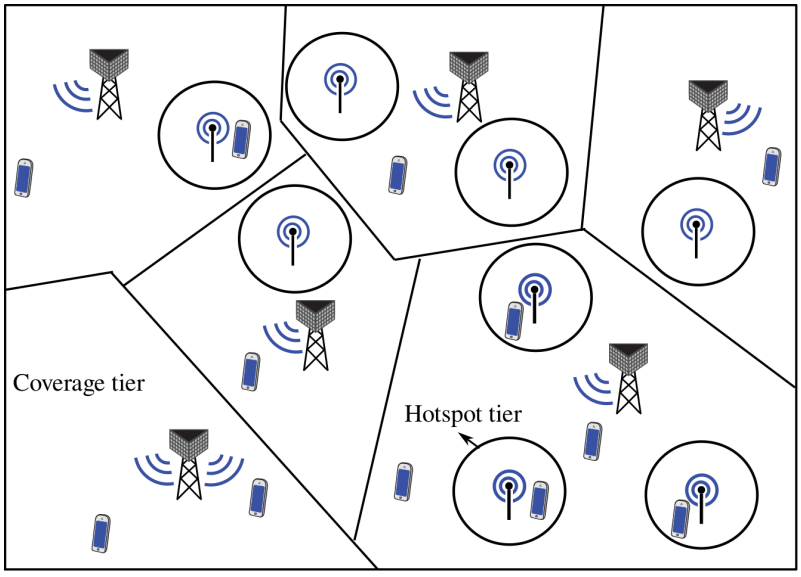


Figure 2.1: Heterogeneous cellular network.

A relevant performance metric of cellular networks is the area throughput. Network densification, increasing the bandwidth, and techniques that improve spectral efficiency (SE) are fundamental ways to increase area throughput and meet the requirements of B5G networks. Solutions that seek to increase the amount of transmitted data without increasing power and bandwidth are desirable to achieve high SE and energy efficiency (EE). For instance, the massive MIMO technology used in fifth-generation (5G) networks employs a large number of antennas to serve a smaller number of UE devices, at least four times as many, to improve the SE and, consequently, the area throughput [20]. The reason for this is that the massive MIMO regime can offer the three well-known advantages of using multiple antennas at extraordinary levels: beamforming, multiplexing, and diversity gains.

The beamforming gain allows the total energy of the signal to be directed in a desired direction. The more antennas there are, the narrower the beam, the greater the signal strength and the SE of the UE. Spatial multiplexing is the ability to focus a transmission on a single point to serve multiple UE devices on the same time–frequency resources and separate them in the spatial domain. Furthermore, as the signal passes through several different, possibly independent channels, it is subject to many changes. By using appropriate signal processing techniques, this can be used to achieve spatial diversity gains, increasing the quality of the signal at the receiver, as well as the robustness and reliability of the transmission link [16].

In this context, two essential propagation phenomena can be defined in massive MIMO systems: channel hardening and favorable propagation. The former means that fading channels behave almost as deterministic channels if the antenna signals are processed properly to neutralize small-scale fading. In principle, the processing makes use of the massive spatial diversity achieved by having many antennas. Favorable propagation means that the channels of spatially separated UE devices are nearly orthogonal in the spatial domain, because transmission and reception are very spatially directive [15].

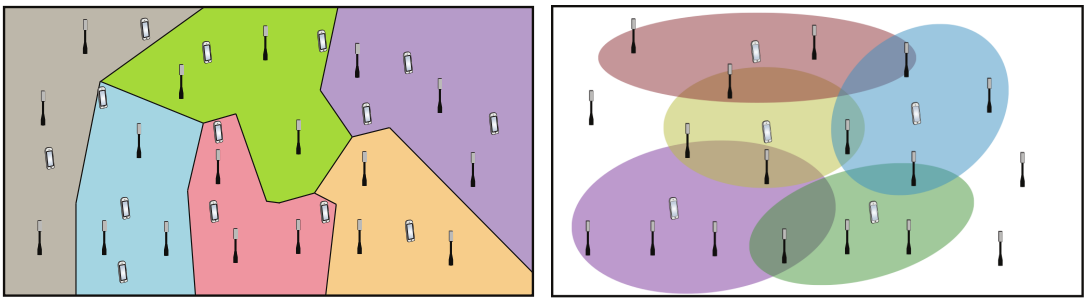
Despite the gains of massive MIMO technology, there are still significant data rate variations in each cell. Moreover, the UE at the cell edge is also affected by interference from neighboring BSs, and thus, the signal-to-interference-plus-noise ratio (SINR) can be substantially lower than the signal-to-noise ratio (SNR) at these locations. Therefore, the primary goal of B5G systems should not be to increase the peak rates but to focus on achieving a more uniform rate among the UE across the different locations in the geographical coverage area. This can be achieved by cell-free systems [15].

2.2 Cell-free massive MIMO networks

CF massive MIMO systems consist of a large number of BSs, here called APs, that are geographically distributed in the coverage area and operate jointly, synchronously, and coherently to serve the UE in the same time–frequency resources. The APs can be coor-

dinated by one or more CPUs through a fronthaul network. The term “cell-free” means that no cell boundaries exist from the perspective of the UE during UL and DL transmission. This can effectively mitigate the inherent problems of the cellular topology, such as interference between cells and a long distance from a BS, which are the main factors for the poor data rate performance of UE at the cell edges.

Although the cell-free concept is receiving great attention from the scientific community as one of the solutions for B5G systems, it presents as an evolution of the similar concept called coordinated multipoint (CoMP), network MIMO, and multicell processing found in previous works and implemented in various systems based on current wireless access technologies. These systems also aim to reduce intercell interference by jointly coordinating signals from different APs. However, due to the conventional network-centric implementation of CoMP, the advantages in terms of spectral efficiency have not been significant in practice. Network-centric operation refers to the case where a subset of APs are divided into disjoint clusters, and the APs in a cluster transmit jointly to the UE residing in their joint coverage area, as illustrated in Figure 2.2a. This essentially creates cells with distributed APs inside, and does not solve the problem of interference among different cells. Another option is to implement CoMP transmission using user-centric clustering, where each UE device selects a set of preferred APs and all the APs that affect the UE take its interference into consideration, as illustrated in Figure 2.2b. It can be noted that user-centric clustering will guarantee the interference control irrespective of the location of the UE devices [15].



(a) Network-centric (b) User-centric

Figure 2.2: (a) Network-centric implementation of CoMP in a cellular network. (b) User-centric implementation of CoMP in a cellular network [15].

Cell-free massive MIMO systems use a combination of previous solutions, such as massive MIMO operation, a dense distributed network topology, and user-centric transmission. In other words, taking the best aspects from the three technologies, they are combined into a single network and then jointly optimized to achieve an ultimate embodiment of a wireless network [15].

2.2.1 System models

Let us assume a CF massive MIMO system consisting of L APs, equipped with N antennas each, and K single-antenna UE devices. The total number of antennas in the network is $M = NL$, where $M > K$. The APs are connected to a CPU through fronthaul links that are assumed to be error-free and able to support the data traffic, as shown in Figure 2.3 [20].

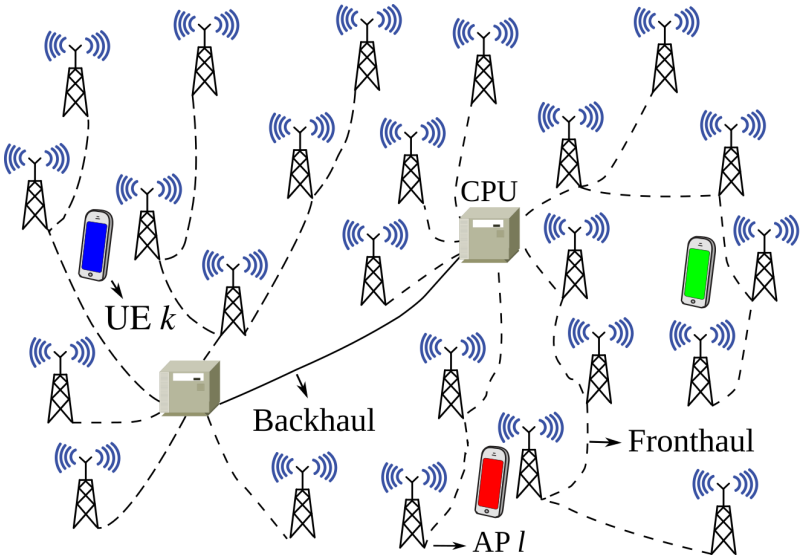


Figure 2.3: Cell-free massive MIMO network.

The system operates using a time-division duplex (TDD) protocol whose frame length equals the coherence interval, which means that the channel is static within a frame but varies independently from frame to frame. In the TDD protocol, UL and DL transmission occur in the same frame. Then, the specific channel realization inside a frame is the same for UL and DL, being sufficient to transmit the pilot only in the UL owing to the channel reciprocity. Based on the received UL signals, each AP can then estimate the channels between itself and all the UE devices. Each coherence block comprises τ_c samples, where τ_{up} samples are reserved for UL pilot signals and τ_d for DL data transmissions, as illustrated in Figure 2.4.

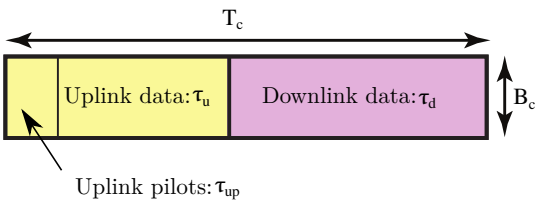


Figure 2.4: TDD protocol under consideration, where each coherence block is used for both UL and DL transmissions [15].

The coherence interval length is given by $\tau_c = T_c B_c$, where T_c and B_c denote the coherence time and bandwidth, respectively. The exact value is difficult to model, but the coherence

time depends on the transmission wavelength (λ) and the UE velocity (v), and the coherence bandwidth can be related to the delay spread (T_d) so that the following rule of thumb can be used to estimate their values: $T_\mathrm{c} = \lambda/4v$ and $B_\mathrm{c} = 1/2T_\mathrm{d}$ [20].

2.2.2 Channel models

The channel vector $\boldsymbol{h}_{kl} \in \mathbb{C}^{N\times 1}$ between AP l and UE k (and vice versa) undergoes independent correlated Rayleigh fading, defined as

$$\boldsymbol{h}_{kl} \sim \mathscr{N}_{\mathbb{C}}\left(\mathbf{0}_N, \boldsymbol{R}_{kl}\right), \tag{2.1}$$

where $\boldsymbol{R}_{kl} \in \mathbb{C}^{N\times N}$ is the covariance matrix, which models the large-scale fading behavior, considering spatial channel correlation, path loss, and shadowing [15], [20].

It is important to note that there are different models for the channel between an AP and the UE, each valid for different propagation assumptions. Rayleigh fading is used to model the small-scale fading characteristics of rich scattering scenarios, i.e., when a large number of non-line-of-sight (NLoS) propagation components exist. However, Rayleigh fading models assume that the line-of-sight (LoS) is blocked, which might not always be true for CF systems. Because CF systems allow the APs to be closer to the UE, a more appropriate model would be to consider the probability that both LoS and NLoS propagation components exist. For this purpose, the Rician fading model can be used. Considering that the channel vector $\boldsymbol{h}_{kl} \in \mathbb{C}^{N\times 1}$ between AP l and UE k undergoes independent correlated Rician fading, $\boldsymbol{h}_{kl}$ can be defined as

$$\boldsymbol{h}_{kl} = \overline{\boldsymbol{h}}_{kl}\, e^{j\psi_{kl}} + \widetilde{\boldsymbol{h}}_{kl}, \tag{2.2}$$

where $\overline{\boldsymbol{h}}_{kl}\, e^{j\psi_{kl}} \in \mathbb{C}^{N\times 1}$ corresponds to the deterministic LoS component, and $\widetilde{\boldsymbol{h}}_{kl} \in \mathbb{C}^{N\times 1}$ is the small-scale fading random component from NLoS propagation. Due to the UE mobility, random phase shifts $\psi_{kl} \sim \mathscr{U}[0, 2\pi)$ occur in the LoS component, which are assumed to be unknown. Assuming that the APs are equipped with a half-wavelength-spacing uniform linear array (ULA), the term $\overline{\boldsymbol{h}}_{kl}$ can be expressed as

$$\overline{\boldsymbol{h}}_{kl} = \sqrt{\frac{\kappa_{kl}}{1+\kappa_{kl}}}\boldsymbol{h}_{kl}^{\mathrm{LoS}} \tag{2.3}$$

with

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

where κ_{kl} is the Rician factor, φ_{kl} is the azimuth angle, θ_{kl} is the elevation angle of the LoS component, and β_{kl} is the large-scale fading, including path loss and shadowing.

The random component can be written as

$$\widetilde{\boldsymbol{h}}_{kl} = \sqrt{\frac{1}{1+\kappa_{kl}}}\boldsymbol{h}_{kl}^{\mathrm{NLoS}} \tag{2.5}$$

where $\boldsymbol{h}_{kl}^{\mathrm{NLoS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \boldsymbol{R}_{kl}^{\mathrm{NLoS}})$, with $\boldsymbol{R}_{kl}^{\mathrm{NLoS}}$ representing the spatial correlation matrix, which is subjected to $\mathrm{tr}\{\boldsymbol{R}_{kl}^{\mathrm{NLoS}}\}/N = \beta_{kl}$. Then, the covariance matrix $\widetilde{\boldsymbol{R}}_{kl} = \mathbb{E}\{\widetilde{\boldsymbol{h}}_{kl}\widetilde{\boldsymbol{h}}_{kl}^{\mathrm{H}}\} \in \mathbb{C}^{N\times N}$ can be expressed as

$$\widetilde{\boldsymbol{R}}_{kl} = \frac{1}{\kappa_{kl}+1}\boldsymbol{R}_{kl}^{\mathrm{NLoS}}. \tag{2.6}$$

The Rician factor κ_{kl} represents the power ratio between the LoS and NLoS components, being generally a function of the distance d_{lk} between AP l and UE k [21]–[24]. For LoS propagation links, the Rician factor can be modeled as $\kappa_{kl} = p_{\mathrm{LoS}}(d_{lk})/(1-p_{\mathrm{LoS}}(d_{lk}))$, where p_{LoS} is the LoS propagation probability [22]. For NLoS propagation links, the Rician factor is equal to zero [24].

2.2.3 Scalability issues and AP selection

The canonical CF massive MIMO system, where all APs serve each UE device, is unfeasible because the processing and resource requirements scale with the number of UE devices. It is also unnecessary in a geographically large network, because the signal strength is not suitable for APs physically distant to the UE [15]. To achieve a scalable CF massive MIMO system, the complexity and resource requirements for each AP must remain finite when the number of UE devices goes to infinity. In this way, the following tasks cannot depend on the number of UE devices: (i) signal processing for channel estimation; (ii) signal processing for data reception and transmission; (iii) fronthaul signaling for data and CSI sharing; and (iv) power control optimization.

One way to achieve a scalable CF massive MIMO is to use a UC approach to form the AP clusters serving each UE device, which can be implemented by following the dynamic cooperation clustering (DCC) framework [25]. The DCC framework enables "unified analysis of anything from interference channels to network MIMO" [25]. This is achieved by defining a set of diagonal matrices that determine which AP antennas may transmit to which UE device [13], [25]. However, to address the scalability issues, it is also necessary to limit the number of UE devices that each AP can serve. Figure 2.5 illustrates how the subset of APs serving each UE device may form, creating overlapping AP clusters.

In UC CF massive MIMO systems, only a subset of APs serves UE k, which is denoted as $\mathscr{M}_k \subset \{1,\ldots,L\}$. It can also be defined as the subset $\mathscr{D}_l$ containing the UE devices served by the AP l. A diagonal matrix $\boldsymbol{D}_{kl} \in \mathbb{N}^{N\times N}$ is employed to represent which APs

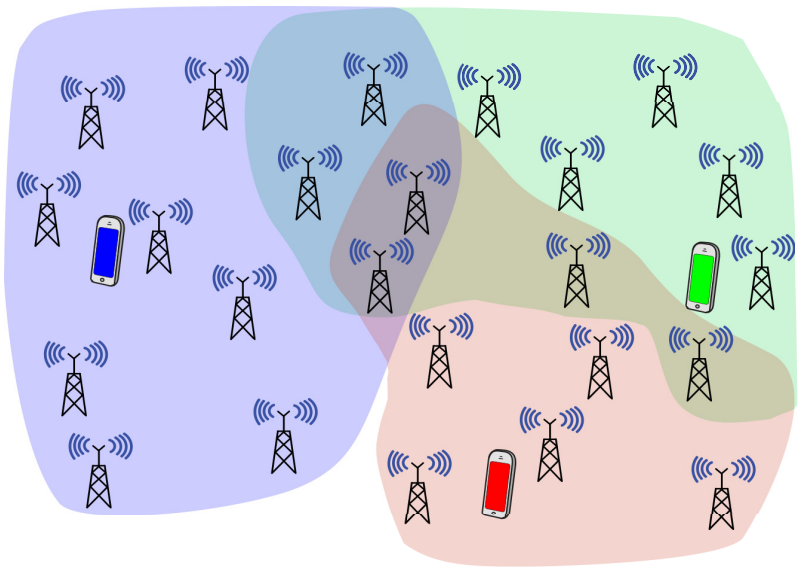


Figure 2.5: UC cell-free massive MIMO network.

are serving UE k, being defined as

$$\boldsymbol{D}_{kl} = \begin{cases} \boldsymbol{I}_N & l \in \mathscr{M}_k \\ \boldsymbol{0}_N & l \notin \mathscr{M}_k. \end{cases} \tag{2.7}$$

One way to evaluate the impact of DCC is to perform the scalable clustering method proposed in [15], where each AP serves τ_{up} UE devices with the best channel conditions to it. Thus, it follows that this strategy yields $|\mathscr{D}_l| = \tau_{up}$. The method uses a joint pilot assignment and AP selection scheme [15, Sec. 4.4]. In this method, the first τ_{up} UE devices are assigned mutually orthogonal pilots, while the remaining UE devices are assigned to the pilot that causes the lowest pilot contamination. After that, each AP chooses to serve τ_{up} UE devices with the largest channel gain in each pilot. This method takes into account both the pilot contamination and scalability issues as the complexity and resource requirements for each AP remain finite when the number of UE devices goes to infinity. Although it is not optimal, the SE performance loss compared with the case where all UE devices are served by all APs is small [15].

2.2.4 UL training

During the UL training phase, all UE devices simultaneously send pilot sequences of τ_{up} length to the APs. Then, the UL channels are estimated by correlating the received signals with a known pilot sequence and performing minimum mean square error (MMSE) estimation. The pilot signals are pairwise orthogonal, and different UE devices can be assigned to the same pilot sequence when $K > \tau_{up}$. Let $t_k \in \{1, \ldots, \tau_{up}\}$ denote the index

of the pilot assigned to UE k, the subset of UE that uses the same pilot as UE k is denoted as $\mathscr{P}_k \subset \{1,\ldots,K\}$, including itself. The correlated received pilot signal at AP l can be expressed as [15]

$$\boldsymbol{y}_{t_k l}^{\text{pilot}} = \sum_{i \in \mathscr{P}_k} \sqrt{\tau_{up}\eta_i}\,\boldsymbol{h}_{il} + \boldsymbol{n}_{t_k l}, \tag{2.8}$$

where η_i represents the UL transmit power of UE i, and $\boldsymbol{n}_{t_k l} \sim \mathscr{N}_{\mathbb{C}}\left(\boldsymbol{0}_N, \sigma_{ul}^2 \boldsymbol{I}_N\right)$ stands for the noise term. The MMSE channel estimate of $\boldsymbol{h}_{kl}$ based on $\boldsymbol{y}_{t_k l}^{\text{pilot}}$ is given by

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

where $\boldsymbol{\Psi}_{t_k l} = \mathbb{E}\{(\boldsymbol{y}_{t_k l}^{\text{pilot}})(\boldsymbol{y}_{t_k l}^{\text{pilot}})^{\text{H}}\} = \sum_{i \in \mathscr{P}_k} \eta_i \tau_{up} \boldsymbol{R}_{il} + \sigma_{ul}^2 \boldsymbol{I}_N$ is the correlation matrix of the received signal in (2.8).

For Rician fading models, the same procedures can be performed to estimate channels, using the phase-unaware linear MMSE estimation, because the random phase shifts ψ_{kl} are assumed to be unknown. The channel estimates of $\boldsymbol{h}_{kl}$ are still given by (2.9), but $\boldsymbol{R}_{kl} = \mathbb{E}\{\boldsymbol{h}_{kl}\boldsymbol{h}_{kl}^{\text{H}}\} = (\overline{\boldsymbol{h}}_{kl}\overline{\boldsymbol{h}}_{kl}^{\text{H}} + \widetilde{\boldsymbol{R}}_{kl})$ and $\boldsymbol{\Psi}_{t_k l} = \mathbb{E}\{(\boldsymbol{y}_{t_k l}^{\text{pilot}})(\boldsymbol{y}_{t_k l}^{\text{pilot}})^{\text{H}}\} = \sum_{i \in \mathscr{P}_k} \eta_i \tau_{up} (\overline{\boldsymbol{h}}_{il}\overline{\boldsymbol{h}}_{il}^{\text{H}} + \widetilde{\boldsymbol{R}}_{il}) + \sigma_{ul}^2 \boldsymbol{I}_N$.

2.2.5 DL data transmission

In the DL mode, each AP serves the UE implementing power control and precoding based on the UL estimates. Let $q_k(n)$ be the nth symbol intended for UE k. It is assumed that $\mathbb{E}\left\{\mathbf{q}(n)\mathbf{q}(n)^H\right\} = \boldsymbol{I}_K$, where $\mathbf{q}(n) \triangleq [q_1(n), \cdots, q_K(n)]^T$. Thus, the data signal sent by AP l can be written as

$$\boldsymbol{x}_l(n) = \sum_{k=1}^{K} \boldsymbol{D}_{kl}\boldsymbol{w}_{kl}q_k(n), \tag{2.10}$$

where $\boldsymbol{w}_{kl} \in \mathbb{C}^{N \times 1}$ represents the precoding vector, such that $\mathbb{E}\left\{\|\boldsymbol{w}_{kl}\|^2\right\} = \rho_{kl}$, with ρ_{kl} being the transmit power that AP l assigns to UE k. After the signal propagates through the channel, UE k receives a linear combination of the signals transmitted by the APs, i.e.,

$$y_{d,k}(n) = \sum_{l=1}^{L} \boldsymbol{h}_{kl}^{\text{H}}\boldsymbol{x}_l(n) + n_{d,k}(n) = \underbrace{\alpha_{kk}q_k(n)}_{\text{desired signal}} + \underbrace{\sum_{k' \neq k}^{K} \alpha_{kk'}q_{k'}(n)}_{\text{inter-user interference}} + \underbrace{n_{d,k}(n)}_{\text{noise}}, \tag{2.11}$$

where

$$\alpha_{kk'} = \sum_{l=1}^{L} \boldsymbol{h}_{kl}^{\text{H}}\boldsymbol{D}_{k'l}\boldsymbol{w}_{k'l}, \quad k' = 1, \cdots, K. \tag{2.12}$$

The noise at the receiver is formulated as $n_{d,k}(n) \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{dl}^2)$. In addition, α_{kk} is called the effective channel for UE k, and $\alpha_{kk'}, k' \neq k$ is the effective interfering channel. They are essentially the summation of the inner product between the channel vector $\boldsymbol{h}_{kl}$ and the precoding vector $\boldsymbol{D}_{k'l}\boldsymbol{w}_{k'l}$ from all the APs.

2.2.6 Combining and precoding vectors

For UL and DL data transmission, the APs use the channel estimates to compute the combining and precoding vectors, respectively. It is important to note that CF systems may use different forms of coordinate transmissions. The most common ones are the fully distributed and fully centralized configurations, shown in Figure 2.6. Put simply, in the distributed configuration, each AP performs the channel estimation, combining/precoding and power allocation locally, while in the centralized configuration these tasks are performed by the CPU [13].

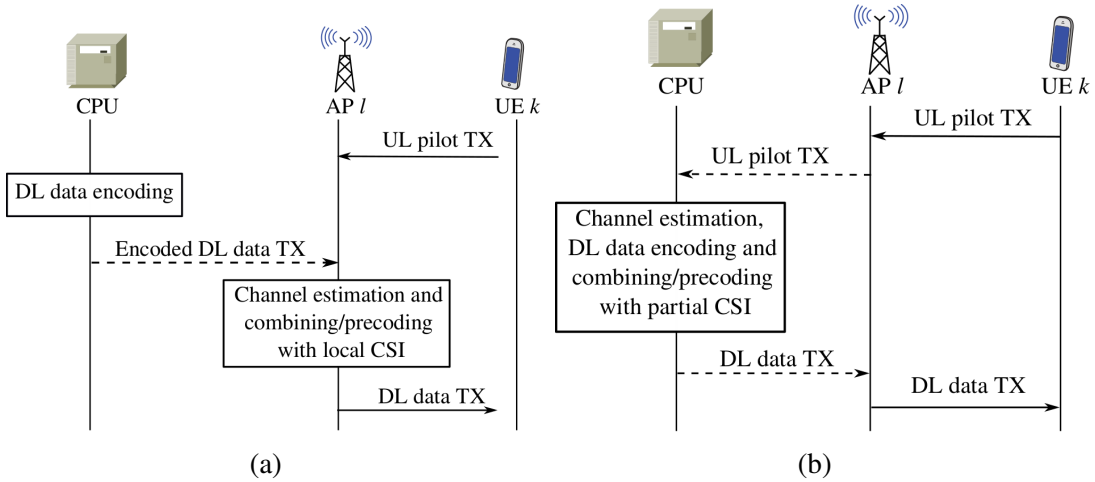


Figure 2.6: Two network implementations under consideration: (a) centralized and (b) distributed processing. Partial CSI means that the CPU can only partially compute the channel estimates because it is based on the received pilot signals of only the APs that serve the UE.

In the DL distributed implementation, illustrated in Figure 2.6a, the only task performed at the CPUs is the encoding of the DL data signals [14]. In this implementation, the network may require less signaling on the fronthaul/backhaul links depending on the system parameters. However, if $\tau_c/(\tau_c - \tau_p) \approx 1$ and $K \gg N$, the distributed implementation may require much more signaling [14].

In the centralized processing, illustrated in Figure 2.6b, AP l sends the received pilot signals to the CPUs during the training phase [14]. Thus, the CPU takes the responsibility to estimate the channels, generate the combining and precoding vectors, and process the DL signals. In this implementation, CPUs have access to channel statistics, provide better

interference cancellation, and reach a higher SE.

Different combining and precoding schemes can be derived for each network implementation. It is noteworthy that MMSE combining is a signal processing technique that minimizes the mean square error (MSE) of data detection and maximizes the SINR for all UE devices in the network. It is based on channel estimates available on the CPU, efficiently suppressing interference among them. MMSE combining can be written as

$$\boldsymbol{v}_k^{\text{MMSE}} = p_k \left(\sum_{i=1}^{K} p_i \boldsymbol{D}_k \left(\widehat{\boldsymbol{h}}_i \widehat{\boldsymbol{h}}_i^{\text{H}} + \boldsymbol{C}_i \right) \boldsymbol{D}_k + \sigma_{ul}^2 \boldsymbol{I}_M \right)^{-1} \boldsymbol{D}_k \widehat{\boldsymbol{h}}_k, \tag{2.13}$$

where $\boldsymbol{C}_k$ is the correlation matrix of the estimation error $\boldsymbol{e}_k = \boldsymbol{h}_k - \widehat{\boldsymbol{h}}_k$, such that $\boldsymbol{C}_k = \mathbb{E}\{\boldsymbol{e}_k \boldsymbol{e}_k^{\text{H}}\}$. For centralized configuration, the signal processing is derived using the collective channel vector, given by $\boldsymbol{h}_k = \left[\boldsymbol{h}_{k1}^{\text{T}}, ..., \boldsymbol{h}_{kL}^{\text{T}}\right]^{\text{T}} \in \mathbb{C}^{M \times 1}$. The distributed version of MMSE is called local MMSE (L-MMSE) combining, which uses the local channel estimates of all UE devices available at each AP, given by

$$\boldsymbol{v}_{kl}^{\text{L-MMSE}} = p_k \left(\sum_{i=1}^{K} p_i \left(\widehat{\boldsymbol{h}}_{il} \widehat{\boldsymbol{h}}_{il}^{\text{H}} + \boldsymbol{C}_{il} \right) + \sigma_{ul}^2 \boldsymbol{I}_N \right)^{-1} \boldsymbol{D}_{kl} \widehat{\boldsymbol{h}}_{kl}. \tag{2.14}$$

These methods are not scalable because they need channel estimates of all UE devices; however, they can be modified to meet the scalability requirements. In the following, some scalable schemes are presented for each network implementation.

To solve the scalability issues, the centralized partial MMSE (P-MMSE) considers only the UE devices that interfere most with UE k. Those UE devices are typically located in the neighborhood of UE k, which corresponds to a small subset of the other UE devices. There may be different approaches to define this subset; one of them is to consider the UE devices that are partially served by the same APs as UE k, given by

$$\mathscr{S}_k = \{ i : \boldsymbol{D}_k \boldsymbol{D}_i \neq \boldsymbol{0}_{M \times M} \}, \tag{2.15}$$

where $\boldsymbol{D}_k = \text{diag}\left(\boldsymbol{D}_{k1}, ..., \boldsymbol{D}_{kL}\right) \in \mathbb{C}^{M \times M}$ is the diagonal block clustering matrix.

Therefore, the P-MMSE combining can be written as [15]

$$\boldsymbol{v}_k^{\text{P-MMSE}} = p_k \left(\sum_{i \in \mathscr{S}_k} p_i \boldsymbol{D}_k \widehat{\boldsymbol{h}}_i \widehat{\boldsymbol{h}}_i^{\text{H}} \boldsymbol{D}_k + \boldsymbol{Z}_{\mathscr{S}_k} + \sigma_{ul}^2 \boldsymbol{I}_M \right)^{-1} \boldsymbol{D}_k \widehat{\boldsymbol{h}}_k \tag{2.16}$$

with

$$\boldsymbol{Z}_{\mathscr{S}_k} = \sum_{i \in \mathscr{S}_k} p_i \boldsymbol{D}_k \boldsymbol{C}_i \boldsymbol{D}_k. \tag{2.17}$$

The partial regularized zero-forcing (P-RZF) simplifies the P-MMSE combining to reduce the computational complexity. It neglects the matrix $\boldsymbol{Z}_{\mathscr{S}_k}$ by noticing that if the channel conditions of the interfering UE devices in $\mathscr{S}_k$ are good, all the corresponding estimation error correlation matrices $\boldsymbol{C}_i : i \in \mathscr{S}_k$ in $\boldsymbol{Z}_{\mathscr{S}_k}$ will be small. This change has a negligible impact on complexity, but it enables further reformulation. Then, the P-RZF combining vector can be written as [15]

$$\boldsymbol{v}_k^{\mathrm{P-RZF}} = \left[\boldsymbol{D}_k \widehat{\boldsymbol{H}}_{\mathscr{S}_k} \left(\widehat{\boldsymbol{H}}_{\mathscr{S}_k}^{\mathrm{H}} \boldsymbol{D}_k \widehat{\boldsymbol{H}}_{\mathscr{S}_k} + \sigma_{ul}^2 \boldsymbol{P}_{\mathscr{S}_k}^{-1} \right)^{-1} \right]_{:,1} \tag{2.18}$$

where $[\cdot]_{:,1}$ denotes the operation of only keeping the first column of its matrix argument, $\widehat{\boldsymbol{H}}_{\mathscr{S}_k} \in \mathbb{C}^{M \times |\mathscr{S}_k|}$ contains the stacked vectors $\widehat{\boldsymbol{h}}_i$ with indices $i \in \mathscr{S}_k$, with the first column being $\widehat{\boldsymbol{h}}_k$, and $\boldsymbol{P}_{\mathscr{S}_k} \in \mathbb{R}^{|\mathscr{S}_k| \times |\mathscr{S}_k|}$ is a diagonal matrix containing the transmit powers p_i for $i \in \mathscr{S}_k$, listed in the same order as the columns $\widehat{\boldsymbol{H}}_{\mathscr{S}_k}$. Essentially, P-RZF combining takes the pseudo-inverse of the partial channel $\widehat{\boldsymbol{H}}_{\mathscr{S}_k}$ and regularizes it by adding $\sigma_{ul}^2 \boldsymbol{P}_{\mathscr{S}_k}^{-1}$ containing the transmit and noise powers. Combining vectors based on pseudo-inverses forces the interference between the UE devices to zero, but it may lead to large losses in the desired signal power when the UE devices have similar channels [15].

To comply with scalability requirements in the distributed implementation, it is possible to consider only the channels of the UE devices that AP l serves in the L-MMSE combining in (2.14), i.e., $k \in \mathscr{D}_l$, because $|\mathscr{D}_l|$ is limited. Therefore, the local partial MMSE (LP-MMSE) combining can be written as

$$\boldsymbol{v}_{kl}^{\mathrm{LP-MMSE}} = p_k \left(\sum_{i \in \mathscr{D}_l} p_i \left(\widehat{\boldsymbol{h}}_{il} \widehat{\boldsymbol{h}}_{il}^{\mathrm{H}} + \boldsymbol{C}_{il} \right) + \sigma_{ul}^2 \boldsymbol{I}_N \right)^{-1} \boldsymbol{D}_{kl} \widehat{\boldsymbol{h}}_{kl}. \tag{2.19}$$

The most common distributed combining is called the maximum ratio (MR) (or conjugated beamforming), which is written as

$$\boldsymbol{v}_{kl}^{\mathrm{MR}} = \widehat{\boldsymbol{h}}_{kl}. \tag{2.20}$$

The MR combining is scalable and has a low complexity. Furthermore, it maximizes the desired signal power, but cannot efficiently mitigate interference among UE devices.

Motivated by the UL–DL duality, the DL precoding vectors can be selected based on the UL combiners [13], [15], [20]. This strategy also reduces the computational complexity

for computing the precoding vectors, because it is only needed to normalize the combining vectors, such as

$$\boldsymbol{w}_k = \sqrt{\rho_k}\frac{\boldsymbol{v}_k}{\sqrt{\mathbb{E}\left\{\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{D}_k\boldsymbol{v}_k\right\}}},\ \boldsymbol{w}_{kl} = \sqrt{\rho_{kl}}\frac{\boldsymbol{v}_{kl}}{\sqrt{\mathbb{E}\left\{\boldsymbol{v}_{kl}^{\mathrm{H}}\boldsymbol{D}_{kl}\boldsymbol{v}_{kl}\right\}}}, \tag{2.21}$$

where $\boldsymbol{v}_k$ and $\boldsymbol{v}_{kl}$ are the arbitrary combining vectors used for centralized and distributed configurations, respectively. If the chosen combining vector is scalable, then the precoding vector is also scalable.

2.2.7 Spectral efficiency

In massive MIMO systems, most works assume that the UE devices have only statistical CSI knowledge, i.e., $\mathbb{E}\{\alpha_{kk}\}$. In order to calculate the SE of DL channels assuming only statistical CSI for data detection, an achievable DL SE for UE k can be derived from the received signal presented in (2.11), being written as [15], [20]

$$\mathrm{SE}_k^{\mathrm{sCSI}} = \frac{\tau_d}{\tau_c}\log_2\left(1+\mathrm{SINR}_k^{\mathrm{sCSI}}\right), \tag{2.22}$$

where τ_d/τ_c is the pre-log factor, which is a fraction of samples per coherence block that is used to transmit the DL data, and the term $\mathrm{SINR}_k^{\mathrm{sCSI}}$ denotes the SINR in the DL direction. From (2.11), the $\mathrm{SINR}_k^{\mathrm{sCSI}}$ can be computed as

$$\mathrm{SINR}_k^{\mathrm{sCSI}} = \frac{\left|\mathbb{E}\{\alpha_{kk}\}\right|^2}{\sum_{k\prime=1}^{K}\mathbb{E}\left\{\left|\alpha_{kk\prime}\right|^2\right\} - \left|\mathbb{E}\{\alpha_{kk}\}\right|^2 + \sigma_{dl}^2}. \tag{2.23}$$

Equation (2.22) is also known as the hardening bound, which is commonly used in massive MIMO theory and is valid for any choice of precoding vectors [14], [26]. It can be seen as a lower bound of capacity and, unfortunately, it does not have a closed-form expression when using P-MMSE and LP-MMSE; nevertheless, it can be computed through Monte Carlo simulations. In addition, all expectations presented in (2.23) are related to the channel realizations [15], [20].

The achievable DL SE when UE k is performing some method to estimate the effective channel can be computed based on the use-and-then-forget (UatF) lower bound, which is written as [27], [28]

$$\mathrm{SE}_k^{\mathrm{eCSI}} = \frac{\tau_d}{\tau_c}\log_2\left(1+\mathrm{SINR}_k^{\mathrm{eCSI}}\right), \tag{2.24}$$

with

$$\mathrm{SINR}_k^{\mathrm{eCSI}} = \frac{\left|\mathbb{E}\left\{\frac{\alpha_{kk}}{\hat{\alpha}_{kk}}\right\}\right|^2}{\mathrm{Var}\left\{\frac{\alpha_{kk}}{\hat{\alpha}_{kk}}\right\} + \sum\limits_{k'\neq k}^{K}\mathbb{E}\left\{\left|\frac{\alpha_{kk'}}{\hat{\alpha}_{kk}}\right|^2\right\} + \sigma_{dl}^2\mathbb{E}\left\{\frac{1}{|\hat{\alpha}_{kk}|^2}\right\}} \tag{2.25}$$

where $\hat{\alpha}_{kk}$ is the effective channel estimate. Other approaches could be used to compute the achievable SE, e.g., the methods presented in [28, Sec. III-A] and [27, Sec. V], but the first one is not a rigorous lower bound, and the second one requires complicated numerical computations, unlike the UatF method.

For comparison purposes, an SE expression can also be derived for the theoretical case where the UE devices know their effective channels perfectly, obtained in some genie-aided manner. The SE for the perfect CSI is given by [15], [28]

$$\mathrm{SE}_k^{\mathrm{pCSI}} = \frac{\tau_d}{\tau_c}\mathbb{E}\left\{\log_2\left(1+\mathrm{SINR}_k^{\mathrm{pCSI}}\right)\right\}, \tag{2.26}$$

where

$$\mathrm{SINR}_k^{\mathrm{pCSI}} = \frac{|\alpha_{kk}|^2}{\sum_{\substack{k'=1\\k'\neq k}}^{K}|\alpha_{kk'}|^2+\sigma_{dl}^2}. \tag{2.27}$$

2.2.8 Energy efficiency

The total energy efficiency $\mathrm{EE}_{\mathrm{tot}}$ in bit/Joule is computed as the ratio between the sum throughput in bit/s and the total power consumed in watts (W) [29], [30]. It can be written as

$$\mathrm{EE}_{\mathrm{tot}} = \frac{B\sum_{k=1}^{K}\mathrm{SE}_k}{P_{\mathrm{tot}}}, \tag{2.28}$$

where B is the bandwidth and SE_k is the achievable DL SE of UE k, without specifying CSI acquisition method. The term P_{tot} is modeled as

$$P_{\mathrm{tot}} = \sum_{l=1}^{L}P_l + \sum_{l=1}^{L}P_{\mathrm{fh},l}, \tag{2.29}$$

where P_l is the power consumed by AP l due to the amplifier and circuit power, and $P_{\mathrm{fh},l}$ is the power that the backhaul link connecting the CPU and AP l consumes. The modeling to compute P_l is based on the closed-form expressions derived for MR precoding [30]. However, P_l can be modified to be a function of $\mathbb{E}\left\{\|\boldsymbol{x}_l\|^2\right\}$ in (2.10), which is valid for any precoding scheme, i.e.,

$$P_l = \frac{1}{\gamma_l}\mathbb{E}\left\{\|\boldsymbol{x}_l\|^2\right\} + NP_{\mathrm{tc},l} \tag{2.30}$$

where $0 < \gamma_l \leq 1$ denotes the efficiency of the power amplifier, and $P_{\text{tc},l}$ is the power required of each antenna of AP l to run internal components, such as converters and filters. The power that the fronthaul link consumes is modeled as

$$P_{\text{fh},l} = P_{0,l} + P_{\text{ft},l} R_{\text{ft},l}, \tag{2.31}$$

where $P_{0,l}$ is the fixed power consumption of each backhaul (traffic-independent power), which may depend on the distances between the APs and the CPU and the system topology, $P_{\text{ft},l}$ is the traffic-dependent power (in watt per bit/s), and $R_{\text{ft},l}$ is the data rate in each fronthaul link. Other aspects may also affect the total power consumption, and consequently, EE. These include, for example, the power consumption of signal processing tasks, such as channel estimation and precoding computations at the APs and CPUs [31].

2.2.9 Power control

Two heuristic methods for power allocation are considered to address the scalability aspects in distributed and centralized network implementations. For the distributed processing, a method that divides the power resources proportionally to the large-scale fading gains of each UE device [15], [32] is used, given by

$$\rho_{kl} = \begin{cases} \rho_d \frac{\sqrt{\beta_{kl}}}{\sum_{i \in \mathscr{D}_l} \sqrt{\beta_{il}}} & \text{if } k \in \mathscr{D}_l \\ 0 & \text{otherwise} \end{cases}, \tag{2.32}$$

where $\beta_{kl} = \text{tr}\left(\boldsymbol{R}_{kl}\right)/N$ and ρ_d is the maximum DL transmit power for each AP.

Fractional power allocation is employed for the centralized configuration [15], given by

$$\rho_k = \rho_d \frac{\left(\sqrt{\sum\limits_{l \in \mathscr{M}_k} \beta_{kl}}\right)^{-1} \left(\sqrt{\omega_k}\right)^{-1}}{\max\limits_{\ell \in \mathscr{M}_k} \sum_{i \in \mathscr{D}_\ell} \left(\sqrt{\sum\limits_{l \in \mathscr{M}_i} \beta_{il}}\right)^{-1} \sqrt{\omega_i}} \tag{2.33}$$

with

$$\omega_k = \max_{\ell \in \mathscr{M}_k} \mathbb{E}\left\{ \left\| \bar{\boldsymbol{w}}_{k\ell} \right\|^2 \right\} \tag{2.34}$$

where $\bar{\boldsymbol{w}}_{k\ell} \triangleq \boldsymbol{v}_{k\ell} \in \mathbb{C}^{N \times 1}$ is the proportion of the centralized precoding/combining $\bar{\boldsymbol{w}}_k \triangleq \boldsymbol{v}_k \in \mathbb{C}^{LN \times 1}$ that corresponds to AP ℓ. The precoding scheme determines how this power is distributed between the different APs, but it is typically the closest APs that contribute with the majority of the power. The normalization factor in the denominator is selected to ensure that none of the APs will transmit with more power than the maximum value ρ_d [15]. It is noteworthy that the two heuristic methods are scalable because their complexi-

ties do not increase with K.

2.2.10 Channel hardening and favorable propagation

Channel hardening and favorable propagation are two important propagation phenomena defined in massive MIMO systems. These properties can also be defined for CF massive MIMO, but it may not present them at the same levels as the conventional cellular massive MIMO. One way is to evaluate the level of channel hardening and favorable propagation based on the effective channel α_{kk} and the effective interfering channel $\alpha_{kk'}$ [15].

The value of the effective channel can potentially remain almost constant even if the individual elements of the L channel vectors $\boldsymbol{h}_{k1}, \cdots, \boldsymbol{h}_{kL}$ are changing. More precisely, when the random realizations of the effective channel are close to the mean value (i.e., the variance is small), then approximately the same effective scalar channel will appear in every coherence block, and thus, we can operate the system as if we were communicating over a deterministic channel. However, as the effective channel is the summation of the inner product between the channel vector $\boldsymbol{h}_{kl}$ and the precoding vector $\boldsymbol{D}_{k'l}\boldsymbol{w}_{k'l}$ from all the APs, the different geographical locations of the APs may lead to different large-scale fading coefficients for each AP, which is the main reason why the channel hardening degree may be lower in CF systems than in cellular ones. This fact can be partially compensated for by the power control, but may not necessarily be desirable from the perspective of the end-to-end performance [15]. The channel hardening degree can be computed from the effective channel as

$$\mathrm{CHD}_k = \frac{\mathrm{Var}\{\alpha_{kk}\}}{\left(\mathbb{E}\{\alpha_{kk}\}\right)^2}. \tag{2.35}$$

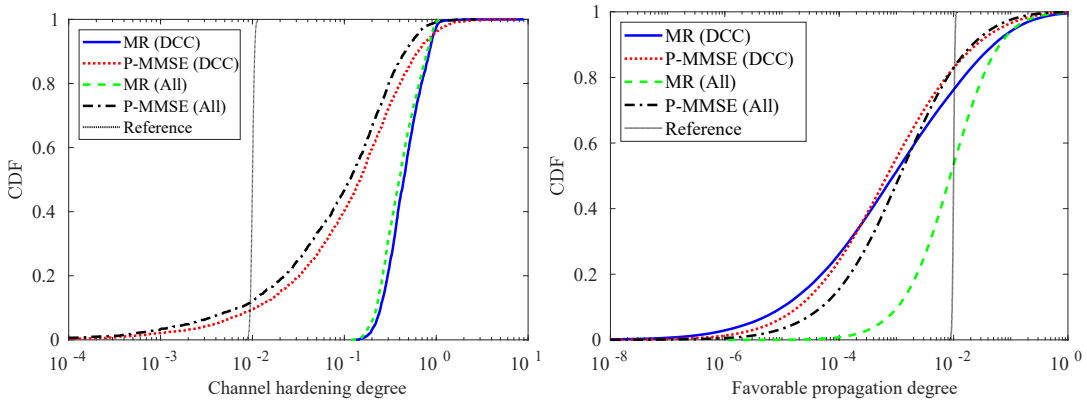
When multiple UE devices are spatially multiplexed in the system, there will generally be interference between their transmissions, and the precoding/combining can be selected to strike a balance between achieving strong desired signals and causing little interference. The favorable propagation degree can be computed from the effective interfering channel by

$$\mathrm{FVP}_k = \frac{\mathrm{Var}\{\alpha_{kk'}\}}{\left(\mathbb{E}\{\alpha_{kk}\}\right)^2},\ k' \neq k. \tag{2.36}$$

Because the MR scheme uses only the channel estimate for each UE device, if the UE channels are spatially orthogonal, then the inter-user interference is automatically mitigated by MR processing. When this happens, we can say that favorable propagation is experienced. However, favorable propagation may not be achieved for UE devices that are partially served by the same APs, requiring the use of precoding methods that are able to actively suppress interference [15].

To provide a more detailed analysis of how the channel hardening and favorable propa-

gation manifest in CF systems, the impact of different precoding schemes, AP selection methods, and channel models is investigated. Figure 2.7 compares the channel hardening and favorable propagation degree in a canonical and UC CF massive MIMO, where "All" denotes the former and "DCC" the latter. One can note from Figure 2.7a that the lower number of UE devices served by each AP in the DCC case degrades the channel hardening compared with canonical CF, but the differences are not significant. Furthermore, MR precoding provides a worse degree of channel hardening than P-MMSE. Thus, it is assumed that the need to estimate the effective channel is higher for MR precoding. Furthermore, compared with the cellular massive MIMO reference case, one cannot rely on hardening in the CF scenario. The opposite effect can be noticed for the favorable propagation in Figure 2.7b, where the DCC framework presents higher degrees of favorable propagation than the canonical CF as fewer simultaneous transmissions are interfering. It can also be observed that MR precoding is more affected than P-MMSE because it is less robust, but their levels of favorable propagation are similar in the case of the DCC approach.

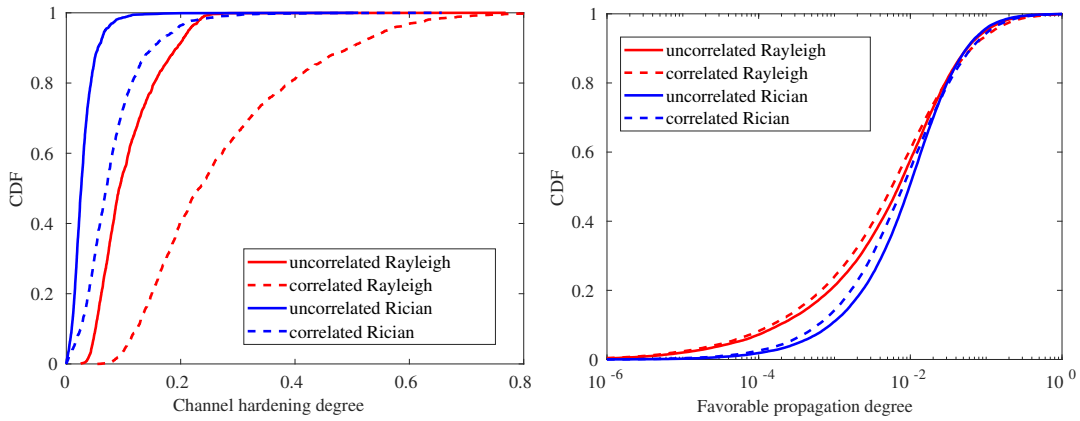


(a) (b)

Figure 2.7: CDF of the channel hardening and favorable propagation in the CF massive MIMO system, computed using (2.35) and (2.36), respectively. The reference case corresponds to a 100-antenna uncorrelated cellular massive MIMO system. The channel hardening and favorable propagation are the better, the closer to zero they are.

It is important to note that there are different models for the channel between an AP and a UE device, each valid for different propagation assumptions. Rayleigh fading is used to model the small-scale fading characteristics of rich scattering scenarios, i.e., when a large number of NLoS propagation components exist. However, the Rayleigh fading model assumes that the LoS is blocked, which may not always be valid for CF systems. Because CF systems allow the APs to be closer to the UE, a more appropriate model would be to consider the probability that both LoS and NLoS propagation components exist. For this purpose, the Rician fading model can be used. Accordingly, Figure 2.8 compares the channel hardening and favorable propagation degrees of different types of channel modeling: correlated or uncorrelated Rayleigh and Rician fading models. The probability

that the LoS component exists in the link between an AP and a UE device is computed following the 3GPP indoor hotspot open office (InH-open) system [33].



(a) (b)

Figure 2.8: CDF of the channel hardening and favorable propagation in the CF massive MIMO indoor system, computed using (2.35) and (2.36), respectively. The scenario corresponds to the 3GPP InH-open system setup assuming a $100\,\text{m} \times 100\,\text{m}$ coverage area, with $L = 100$ APs, each equipped with $N = 4$ antennas, serving $K = 20$ UE devices, and the AP and UE antenna heights are $h_{\text{AP}} = 3$ m and $h_{\text{UE}} = 1.65$ m, respectively. Different channel models are compared using MR precoding: correlated or uncorrelated Rayleigh and Rician fading models.

It can be observed from Figure 2.8a that the spatial correlation of antenna elements has a negative impact on the level of channel hardening for all types of channels as it reduces the spatial diversity gain. Another important observation is that the type of channel modeling also impacts the channel hardening degree. For instance, Rician fading has higher levels of channel hardening than the Rayleigh fading model. This is due to the deterministic LoS component of the Rician model, leading to a weaker random channel variation. This means that scenarios where Rician fading propagation is more common may not benefit as much from the UE devices estimating their effective channels. However, it is important to note that the link LoS probability decreases with the distance, and not all the UE devices may experience it in the CF network. This further motivates the development of effective channel estimation strategies dynamically adapted to the actual needs of each UE device, as the proposed hybrid estimation in Section 3.3. Figure 2.8b shows the impact of the channel model on the favorable propagation degree. It can be noted that the spatial correlation and the presence of the deterministic LoS component have a lower impact on the favorable propagation phenomenon than on the channel hardening. Nonetheless, the favorable propagation degree is slightly worse for the Rician fading model than for the Rayleigh one. This is due to the deterministic LoS component impairing the orthogonality between the channels of different UE devices, depending on their geographical locations, or more specifically, their nominal angles.

3 Overview of publications

This chapter presents a summary of the contributions of Publications I, II, III, IV, V.

For each publication, the relationship between the objectives of this dissertation and the context, motivation, methodology, and contributions of the publication in question is discussed. Table 3.1 shows the relationship between each publication and the research questions outlined in this doctoral dissertation.

Table 3.1: Publications I, II, III, IV, V and their respective research questions.

Publication	**Title**	**RQ**
Publication I	Effective Channel Blind Estimation in Cell-Free Massive MIMO Networks	RQ 1
Publication II	Effective Channel DL Pilot-Based Estimation in User-Centric Cell-Free Massive MIMO Networks	RQ 2
Publication III	Effective Channel Hybrid Estimation in User-Centric Distributed Massive MIMO Networks	RQ 3
Publication IV	Trajectory Optimization in User-Centric Distributed Massive MIMO Systems Enabled by UAV Swarms	RQ 4
Publication V	Scalable User-Centric Distributed Massive MIMO Systems With Restricted Processing Capacity	RQ 5

3.1 Publication I: Effective Channel Blind Estimation in Cell-Free Massive MIMO Networks

3.1.1 Research objectives

This work aims to investigate the impact of relying on the channel hardening property in UC CF massive MIMO networks. Furthermore, the objectives of this research are to identify scenarios where using conventional statistical CSI may be detrimental and to develop an algorithm that can be used to boost the performance in these situations. Specifically, a blind effective channel estimation algorithm for DL data decoding is proposed and evaluated in different case scenarios.

Publication I answers the following research question:

RQ 1: Is it detrimental to rely on statistical CSI for DL data decoding in UC CF massive MIMO systems?

3.1.2 Rationale and context

It is noteworthy that massive MIMO systems operating in rich scattering environments present two important propagation phenomena by taking advantage of spatial diversity

and multiplexing gains: (i) channel hardening, where the channel gain behaves more deterministically even though it is modeled as a random variable, being close to its mean value; and (ii) favorable propagation, where the channels of different UE devices are orthonormal, allowing for directivity in transmission and reception and reduced interference. UC CF massive MIMO relies on these properties, also being present in their deployments.

In Publication I, we briefly introduce the requirements of coherent DL data decoding by emphasizing that the UE must know their effective channel value. Next, we discuss the effect of having specially distributed APs on the channel hardening property, which is conventionally used in massive MIMO systems to estimate the effective channel as its mean value. A need for alternative solutions is identified in cases where the channel hardening property does not hold in UC CF massive MIMO networks, especially those that do not require any extra time–frequency resources to be used for channel estimation. Thus, a blind estimation algorithm is sought, which is detailed next.

3.1.3 Methodology

The blind estimation algorithm leverages the knowledge of statistical asymptotic behavior to estimate the effective channel based on the sample average power of the DL-received signal, i.e., without using any explicit time–frequency samples for channel estimation. The assumption is that the number of DL data samples, UE devices, and transmitting antennas is large, allowing for the effective channel value being clearly identified in the sample average power of the DL-received signal. The UE devices also have to acquire statistical CSI to estimate the value of the effective channel. Figure 3.1 illustrates the steps of the proposed blind estimation algorithm. In the following, a step-by-step proof of the blind estimation algorithm is presented.

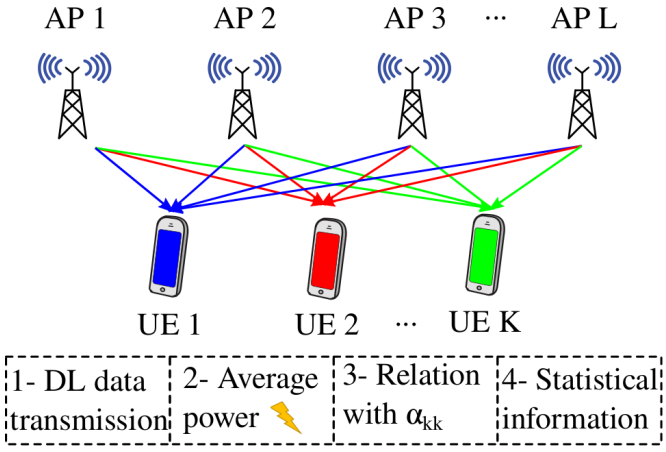


Figure 3.1: Steps of the effective channel blind estimation algorithm.

In order to coherently detect the transmitted data symbol q_k from its received signal in (2.11), UE k needs to know the effective channel α_{kk}, which can be acquired by estimation

methods or relying on statistical CSI [15], [20], [27], [28], [34]. Here, a blind estimation of the effective channel gain α_{kk} is proposed to eliminate the need for any DL pilots. The first step of the blind estimation algorithm is to compute the sample average power of the received signal at UE k per coherence interval, i.e.,

$$\xi_k = \frac{\sum_{n=1}^{\tau_d} \left| y_{d,k}(n) \right|^2}{\tau_d}, \tag{3.1}$$

where $y_{d,k}(n)$ denotes the nth sample received at UE k, shown in (2.11), and $\tau_d = \tau_c - \tau_{up}$ is the number of symbols per coherence interval used for DL data transmission. It is expected that the channel and noise randomness average out in ξ_k when the length of the coherence interval is large. To show this, the first step is to substitute (2.11) in (3.1), which can then be rewritten as

$$\begin{aligned}
\xi_k &= \frac{\sum_{n=1}^{\tau_d} \left| y_{d,k}(n) \right|^2}{\tau_d} \\
&= \frac{1}{\tau_d} \sum_{n=1}^{\tau_d} \Bigg| \underbrace{\alpha_{kk} q_k(n) + \sum_{k' \neq k}^{K} \alpha_{kk'} q_{k'}(n)}_{A_n} + n_{d,k}(n) \Bigg|^2 \\
&= \frac{1}{\tau_d} \sum_{n=1}^{\tau_d} |A_n|^2 + \frac{1}{\tau_d} \sum_{n=1}^{\tau_d} A_n^* n_{d,k}(n) + \frac{1}{\tau_d} \sum_{n=1}^{\tau_d} A_n n_{d,k}^*(n) + \frac{1}{\tau_d} \sum_{n=1}^{\tau_d} |n_{d,k}(n)|^2,
\end{aligned} \tag{3.2}$$

where $(\,)^*$ denotes the conjugate operation. Applying the law of large numbers, when $\tau_d \to \infty$, the second and third terms converge to their zero mean, and the fourth term converges to its mean value σ_{dl}^2. It remains to determine the asymptotic expression for the first term $\frac{1}{\tau_d} \sum_{n=1}^{\tau_d} |A_n|^2$. Because the mean value of $|A_n|^2$ has to be determined, it can be expanded as

$$\begin{aligned}
|A_n|^2 &= \left(\alpha_{kk} q_k(n) + \sum_{k' \neq k}^{K} \alpha_{kk'} q_{k'}(n) \right) \left(\alpha_{kk} q_k(n) + \sum_{k' \neq k}^{K} \alpha_{kk'} q_{k'}(n) \right)^* \\
&= |\alpha_{kk} q_k(n)|^2 + (\alpha_{kk} q_k(n))^* \sum_{k' \neq k}^{K} \alpha_{kk'} q_{k'}(n) \\
&+ \alpha_{kk} q_k(n) \sum_{k' \neq k}^{K} (\alpha_{kk'} q_{k'}(n))^* + \sum_{k' \neq k}^{K} |\alpha_{kk'} q_{k'}(n)|^2.
\end{aligned} \tag{3.3}$$

Because $\mathbb{E}\left\{ \mathbf{q}(n) \mathbf{q}(n)^H \right\} = \boldsymbol{I}_K$, where $\mathbf{q}(n) \triangleq [q_1(n), \cdots, q_K(n)]^T$, all cross-terms between different UE devices have a zero mean and $\mathbb{E}\left\{ |q_k(n)|^2 \right\} = 1$. Therefore, the mean value

of $|A_n|^2$ is given by

$$\mathbb{E}\{|A_n|^2\} = |\alpha_{kk}|^2 + \sum_{k' \neq k}^{K} |\alpha_{kk'}|^2. \tag{3.4}$$

Applying the law of large numbers to (3.2) when $\tau_d \to \infty$, the first term $\frac{1}{\tau_d}\sum_{n=1}^{\tau_d} |A_n|^2$ converges to its mean value in (3.3), the second and third terms converge to zero, and the fourth converges to σ_{dl}^2. Therefore, from the law of large numbers, for $\tau_d \to \infty$, it follows that

$$\xi_k \xrightarrow{P} \left(|\alpha_{kk}|^2 + \sum_{k' \neq k}^{K} |\alpha_{kk'}|^2 + \sigma_{dl}^2 \right). \tag{3.5}$$

There are many unknowns in (3.5) that prevent computing α_{kk} from ξ_k. To solve this issue, we can use once again the law of large numbers to approximate the inter-user interference term $\sum\limits_{k' \neq k}^{K} |\alpha_{kk'}|^2$ by its mean [27]. To this end, (3.5) is first divided by the number of UE devices K

$$\frac{1}{K}\xi_k \simeq \frac{1}{K}|\alpha_{kk}|^2 + \frac{1}{K}\sum_{k' \neq k}^{K} |\alpha_{kk'}|^2 + \frac{1}{K}\sigma_{dl}^2. \tag{3.6}$$

The second term $\frac{1}{K}\sum\limits_{k' \neq k}^{K} |\alpha_{kk'}|^2$ in (3.6) can then be reformulated by adding and subtracting $\sum\limits_{k' \neq k}^{K} \mathbb{E}\left\{ |\alpha_{kk'}|^2 \right\}$. The following asymptotic equivalence can be found

$$\frac{1}{K}\sum_{k' \neq k}^{K} |\alpha_{kk'}|^2 \simeq \frac{1}{K}\sum_{k' \neq k}^{K} \mathbb{E}\left\{ |\alpha_{kk'}|^2 \right\} + \frac{1}{K}\sum_{k' \neq k}^{K} \underbrace{\left(|\alpha_{kk'}|^2 - \mathbb{E}\left\{ |\alpha_{kk'}|^2 \right\} \right)}_{X_{kk'}}. \tag{3.7}$$

Note that $X_{kk'}$ has a zero mean and bounded variance. Thus, applying the law of large numbers, $\frac{1}{K}\sum_{k' \neq k}^{K} X_{kk'} \xrightarrow{P} 0$ in the asymptotic regime, i.e., when $K \to \infty$. Therefore:

$$\frac{1}{K}\sum_{k' \neq k}^{K} |\alpha_{kk'}|^2 - \frac{1}{K}\sum_{k' \neq k}^{K} \mathbb{E}\left\{ |\alpha_{kk'}|^2 \right\} \xrightarrow{P} 0. \tag{3.8}$$

Replacing (3.8) in (3.6), the following asymptotic equivalence can be obtained when $\tau_d, K \to \infty$

$$\frac{1}{K}\xi_k \simeq \frac{1}{K}|\alpha_{kk}|^2 + \frac{1}{K}\sum_{k' \neq k}^{K} \mathbb{E}\left\{ |\alpha_{kk'}|^2 \right\} + \frac{1}{K}\sigma_{dl}^2. \tag{3.9}$$

Although this is an asymptotic result, it is used as motivation to tightly approximate $\sum_{k' \neq k}^{K} |\alpha_{kk'}|^2$ by its mean value. As a consequence, when K and τ_d are large, ξ_k in (3.1)

can be approximated by

$$\xi_k \approx \left(|\alpha_{kk}|^2 + \sum_{k' \neq k}^{K} \mathbb{E}\left\{ |\alpha_{kk'}|^2 \right\} + \sigma_{dl}^2 \right). \tag{3.10}$$

The probability that the real part of α_{kk} is much larger than the imaginary part of α_{kk} increases when the number of APs serving each UE end/or antennas per AP are large, i.e., as the channel hardening improves [15], [27], [28]. Based on this, the effective channel in (3.10) can be further approximated by

$$\alpha_{kk} \approx |\alpha_{kk}| \approx \sqrt{\xi_k - Z_k}, \tag{3.11}$$

where

$$Z_k = \sum_{k' \neq k}^{K} \mathbb{E}\left\{ |\alpha_{kk'}|^2 \right\} + \sigma_{dl}^2. \tag{3.12}$$

Considering the case where the square root argument in (3.11) is nonpositive, the statistical CSI approach that estimates the effective channel by its mean value is used, and the effective channel blind estimator can finally be formulated by

$$\hat{\alpha}_{kk} = \begin{cases} \sqrt{\xi_k - Z_k}, & \text{if } \xi_k > Z_k, \\ \mathbb{E}\{\alpha_{kk}\}, & \text{otherwise.} \end{cases} \tag{3.13}$$

Such an estimator is the same as in [27], but for a UC cell-free massive MIMO system instead of the co-located one. It is assumed that the UE device knows $\mathbb{E}\{\alpha_{kk}\}$ and Z_k because they are functions only of the large-scale coefficients, which are constant over many coherence intervals [15], [27]. Before deriving the achievable DL SE for blind estimation, it can be noted that $\hat{\alpha}_{kk}$ is correlated with the data signal, which makes the derivation of SE too intractable. As in [27], ξ_k in (3.1) is rewritten as

$$\xi_k'(n) = \frac{\sum_{n'=1, n' \neq n}^{\tau_d} \left| y_{d,k}(n') \right|^2}{\tau_d - 1}. \tag{3.14}$$

Thus, the estimate $\hat{\alpha}_{kk}$ is computed using $\xi_k'(n)$ instead of ξ_k in (3.1). For large values of τ_d, it is expected that this modification has a low impact on the channel estimates. Then, the DL SE using blind estimation can be computed using the UatF bound in (2.24).

3.1.4 Results

The results of Publication I focus on comparing the blind estimation algorithm with conventional statistical CSI, which relies on the channel hardening property, and perfect CSI, which is used in a genie-aided manner solely for comparison purposes. In addition, the impact of relaxing the asymptotic assumptions used to derive the blind estimation algo-

rithm is assessed. The main results that answer RQ 1 are reproduced in Figures 3.2 and 3.3, which vary the number of antennas per AP and UE devices, respectively.

Figure 3.2 exhibits the average SE versus the number of antennas per AP. Increasing N improves the channel hardening, and the assumption that the real part of the effective channel is larger than the imaginary becomes true. This result evidences that relying on the channel hardening property of UC CF massive MIMO can be detrimental to the achievable SE performance, especially when using MR precoding. In Figure 3.3, the average normalized mean square error (NMSE) between the effective channel and its estimate is plotted as a function of the number of UE devices K. It is important to note that increasing K improves the approximation of $\sum_{k' \neq k}^{K} |\alpha_{kk'}|^2$ by its mean, which would enhance the performance of the blind estimator. However, as the number of UE devices increases, pilot contamination due to UL pilot reuse has a negative impact on the performance of both statistical CSI and blind estimation. Nonetheless, these results show that the blind estimation algorithm is a good alternative for estimating the effective channel in UC CF massive MIMO because it can improve the SE and the NMSE compared with the statistical CSI approach.

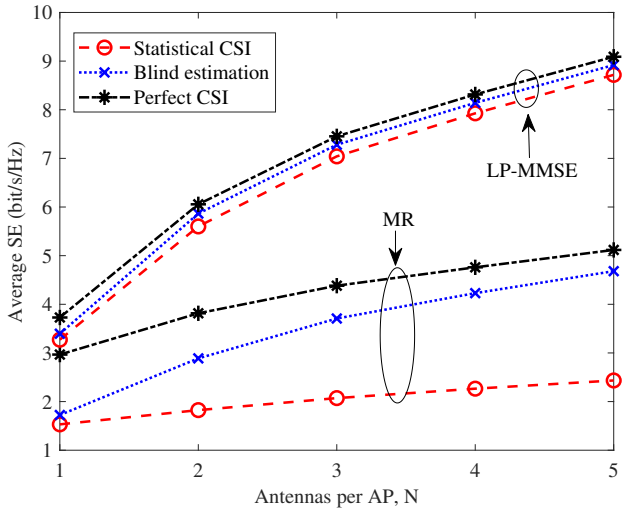


Figure 3.2: Average SE versus the number of antennas per AP, N (Publication I).

3.1.5 Conclusions

Publication I proposed and evaluated a novel blind estimation algorithm for UC CF massive MIMO networks. Furthermore, the results demonstrated that blind estimation outperforms the conventional statistical CSI approach even when the asymptotic assumptions used to derive the algorithm do not hold. Thus, this algorithm can be used as an alternative solution to address the effects of the lack of channel hardening in UC CF massive MIMO networks.

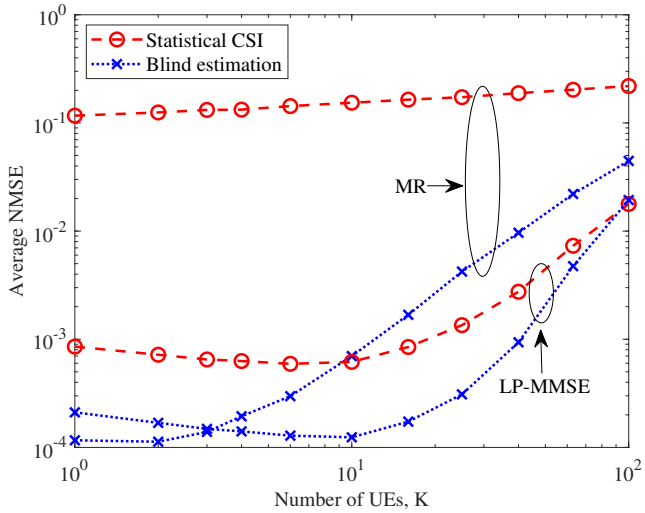


Figure 3.3: Average NMSE between the effective channel and its estimate versus the number of UE devices, K (Publication I).

3.2 Publication II: Effective Channel DL Pilot-Based Estimation in User-Centric Cell-Free Massive MIMO Networks

3.2.1 Research objectives

This work aims to investigate the performance of DL pilot-based training to estimate the effective channel in UC CF massive MIMO networks for different network implementation, precoding schemes, AP selection schemes, and the presence of pilot contamination. Furthermore, we propose a scalable algorithm for DL pilot assignment based on the level of interference between each UE device, aiming to minimize pilot contamination. Finally, Publication II examines whether the DL pilot-based estimation outperforms the blind estimation algorithm proposed in Publication I.

Publication II answers the following research question:

RQ 2: What type of effective channel estimation should be used for DL data decoding in UC CF massive MIMO systems?

3.2.2 Rationale and context

In massive MIMO systems operating under the TDD protocol, the coherence interval length is subdivided into samples for UL pilots, UL data, and DL data. By using the channel reciprocity property, the UL channel estimates are used for precoding the DL data, and the UE relies only on statistical CSI for DL data decoding. This leads to a scenario where there is no need for DL pilots for channel estimation, which reduces the estimation overhead. However, due to the lower level of channel hardening in UC CF massive MIMO

systems, DL effective estimation may be needed. Among different options, it could be included in the TDD protocol τ_{dp} samples used for DL pilot-based estimation, as shown in Figure 3.4. Thus, Publication II investigates the DL pilot-based channel estimation to assist in coherent DL data decoding, proposing proper pilot assignment and assessing whether the performance improvement can compensate for the overhead for estimation.

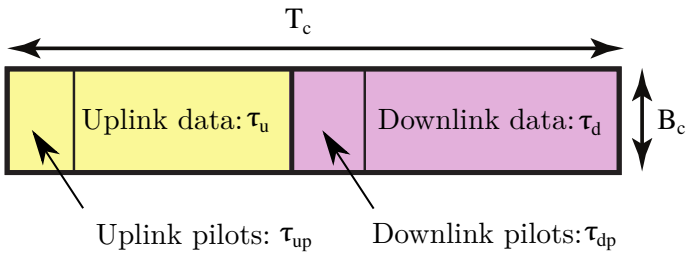


Figure 3.4: TDD protocol under consideration, where each coherence block is used for both UL and DL transmissions, including DL pilot-based training.

3.2.3 Methodology

This work follows the beamforming approach initially proposed in [35], where the APs precode the DL pilots based on the available UL channel estimates for each UE device. This makes the number of resource samples used on orthogonal pilot sequences a function of the number of UE devices instead of the total number of transmit antennas [28], [35]–[37]. It is noteworthy that the beamforming approach is used to estimate only the DL effective channel gain α_{kk}, whereas to estimate the channel vector $\boldsymbol{h}_{kl}$, the number of required pilots is proportional to the transmitting antennas. The main drawback of DL pilot-based estimation is still that it has to use additional resource samples in training pilots, requiring a small number of pilot sequences to keep the overhead—and consequently, their reuse—low among the UE. Therefore, the results of this work consist of assessing the benefits and drawbacks of this DL pilot-based approach compared with other channel estimation technics. Figure 3.5 illustrates the steps of the DL pilot-based algorithm of this work.

The problem of the DL pilot reuse is investigated in [28], where joint UL and DL pilot assignment algorithms are proposed to minimize pilot contamination. The joint UL and DL strategy assigns orthogonal DL pilots to UEs using the same UL pilots. However, the joint strategies require that the number of samples used for UL and DL estimation be greater than or equal to the number of UEs, which means they are not scalable solutions. Therefore, scalable DL pilot assignment solutions are still required to prevent the pilot samples from being dependent on the number of UE devices. In Algorithm 1, a DL pilot assignment method is introduced to minimizes the pilot contamination interference, where the index of the pilot assigned to UE k is denoted as $z_k \in \{1, \ldots, \tau_{dp}\}$. In this case, the first τ_{dp} UE devices are assigned to orthogonal sequences. The remaining UE devices are assigned to the pilot that causes the lowest pilot contamination, given by the average

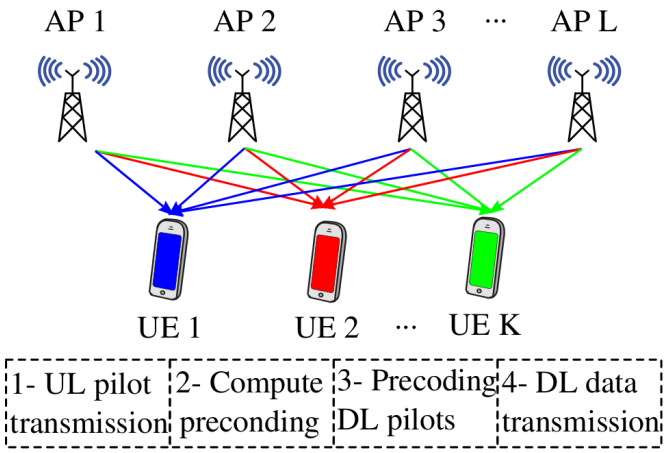


Figure 3.5: Steps of the DL pilot-based training algorithm.

power of the DL pilot contamination term $\mathbb{E}\{|\alpha_{ki}|^2\}$. This method is similar to the UL pilot assignment proposed by [15], with the difference being the pilot contamination term. In this work, UL and DL pilot assignment methods are performed in a disjoint manner; thus, the number of required UL and DL pilot samples is independent of the number of UE devices.

Algorithm 1: DL pilot assignment that minimizes pilot contamination

Input: DL pilot length τ_{dp}.

for $k = 1, \ldots, \tau_{dp}$ **do**
  $z_k \leftarrow k$
end
for $k = \tau_{dp} + 1, \ldots, K$ **do**
  $\zeta \leftarrow \underset{z \in \{1,\ldots,\tau_{dp}\}}{\arg\min} \sum_{l=1, z_l = z}^{k-1} \mathbb{E}\{|\alpha_{ki}|^2\}$
  $z_k \leftarrow \zeta$
end

Output: Pilot assignment indexes $z_1, \ldots, z_K$

3.2.4 Results

The results of Publication II present a comprehensive investigation of the proposed DL pilot assignment and AP selection refinement methods. The results are assessed in terms of SE, EE, NMSE, and CC. The main results that answer RQ 2 are reproduced in Figures 3.6 and 3.7. Figure 3.6 shows the cumulative distribution function (CDF) of the NMSE of different DL pilot assignment methods. We can see that the proposed DL pilot assignment in Algorithm 1, which is a nonjoint method that aims to minimize pilot contamination, performs better than the others, reducing the NMSE by 85% compared with the joint method in the 50$^{\text{th}}$ percentile. This demonstrates that the proposed method effectively mitigates

the impact of pilot contamination on channel estimation, while also having the advantage of being a scalable solution. Having selected the nonjoint minimum pilot contamination as the appropriate DL pilot assignment method, we can fairly compare the DL pilot-based estimation with the other approaches. For that purpose, Figure 3.7 evaluates the SE performance of different effective channel estimation methods by varying the coherence interval length. From this result, it is clear that the best estimation method is different at each value. Nevertheless, the DL pilot-based estimation is the best approach only when the coherence interval is large, because the overhead effect is small in this case.

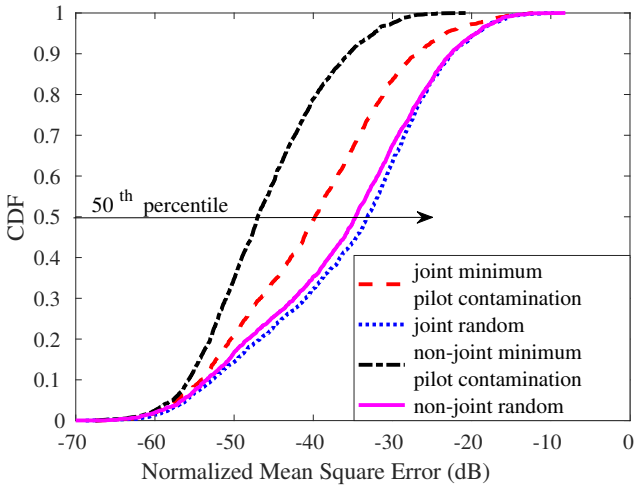


Figure 3.6: CDF of the NMSE for different DL pilot assignment methods using MR precoding (Publication II).

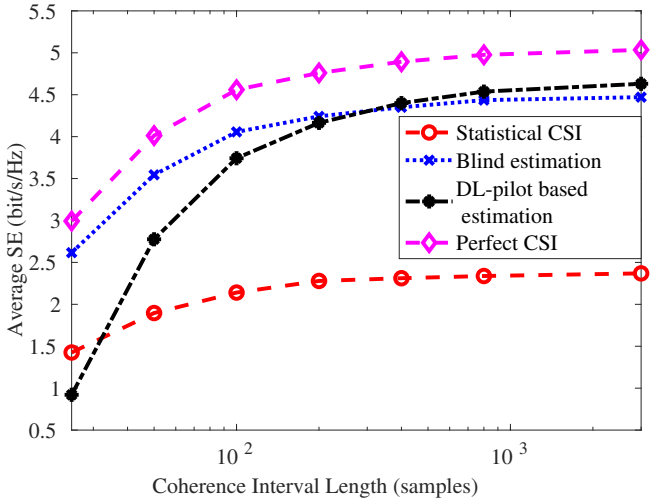


Figure 3.7: Average SE versus coherence interval length (Publication II).

3.2.5 Conclusions

Publication II investigated the performance of the DL pilot-based estimation in UC CF massive MIMO systems. Moreover, an algorithm was proposed for DL pilot assignment that aims to minimize pilot contamination. The results focused on assessing which estimation method performs better: statistical CSI, blind estimation, or DL pilot estimation. Contrary to expectations, despite the fact that the DL pilot-based estimation presents a higher accuracy for channel estimation, the SE performance may be compromised due to the overhead impact. In summary, the best estimation method depends on the system parameters, precoding choice, and the specific performance metric evaluated.

3.3 Publication III: Effective Channel Hybrid Estimation in User-Centric Distributed Massive MIMO Networks

3.3.1 Research objectives

The analyses presented in Publications I and II indicate that any one of the three evaluated methods may be the best choice depending on the characteristics of the UC CF massive MIMO system and the performance metric, which opens avenues for self-regulated resource management strategies. Thus, Publication III aims at proposing a suitable dynamic channel estimation method for each UE, being a hybrid method where it can be used either the blind algorithm, DL pilot-based training, or statistical CSI method.

Publication III answers the following research question:

RQ 3: How can a dynamic channel estimation approach be developed for UC CF massive MIMO networks?

3.3.2 Rationale and context

The previous results presented in Publications I and II assume that only NLoS components are present, modeling the wireless channel as Rayleigh fading. However, in UC CF massive MIMO systems, the APs are generally located close to UE, and therefore, it is important to consider both the LoS and NLoS components, in which case the is modeled as Rician fading. In the case of Rician fading, some UE devices may also experience a higher degree of channel hardening if the LoS component is present. The channel hardening degree is evaluated in Figure 2.8a, where it is also possible to observe a behavior where channel hardening properties vary considerably among the UE. Notably, this behavior is also present in UC CF massive MIMO systems with Rayleigh fading, and the UE may experience different levels of channel hardening. Thus, some UE devices may rely on channel hardening without a significant impact on their performance if they present a high channel hardening degree, while the others do not. This behavior motivates toward a solution that can be applied in a UE base, i.e., where some UE devices can be assigned to a more advanced effective channel estimation method if this solution

can improve their performance. In contrast, other UE devices may rely effectively on channel hardening properly or use blind estimation techniques, saving possible resources that are needed for an advanced channel estimation procedure. This scenario is illustrated in Figure 3.8, where it is also possible to include other aspects such as UE mobility and quality-of-service (QoS), which may influence the need for advanced channel estimation procedures. For instance, UE devices with a high velocity may face channel aging issues and a need to estimate their channel more frequently. Moreover, the possible distinct QoS requirements may be a reason for the use of advanced estimation methods to achieve better performance.

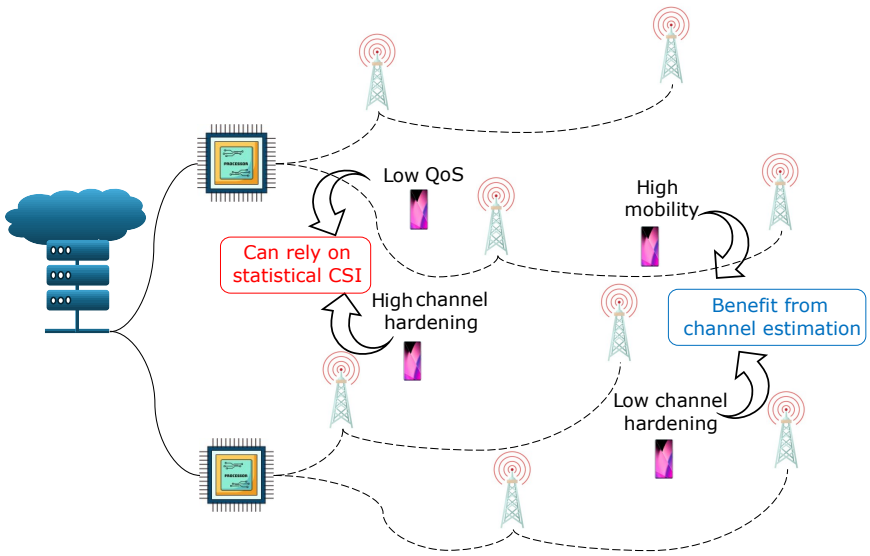


Figure 3.8: UC cell-free massive MIMO scenario for effective channel estimation.

It is worth noting the benefits and drawbacks of statistical CSI, blind estimation, and DL pilot-based estimation in terms of sensitivity, accuracy, computational complexity, and overhead. Statistical CSI and blind estimation are more sensitive to the network configuration and implementation, e.g., their performance depends on the number of antennas per AP and APs serving each UE device. In terms of estimation accuracy, DL pilot estimation is the best option. Still, pilot contamination can negatively impact its performance, which occurs when the number of pilots is lower than the number of UE devices. In such a scenario, the need for proper DL pilot assignment to reduce the pilot contamination effect also leads to a higher computational complexity. Furthermore, the additional time–frequency resources used for DL pilots lead to a higher overhead. Given these benefits and drawbacks of the different effective channel estimation strategies, selecting only one method for all UE devices may not be the best approach. Thus, Publication III proposes a strategy to allow each UE device to estimate its effective channel either with the conventional statistical CSI, relying on the hardening property, or use better estimation strategies, such as blind estimation and DL pilot-based estimation.

3.3.3 Methodology

The benefits and drawbacks of each channel estimation alternative can be assessed in terms of key performance metrics: accuracy based on the NMSE between estimation and the effective channel, and the achievable SE including the overhead in the form of a pre-log factor. Therefore, the proposed decision to define which estimation method each UE device should use relies on these key performance metrics. It is also possible to prioritize the NMSE or the SE by adjusting a specific coefficient. Although the UE mobility and QoS are also important, a decision metric including these aspects is left for future work. Figure 3.9 illustrates the proposed hybrid DL channel estimation strategy.

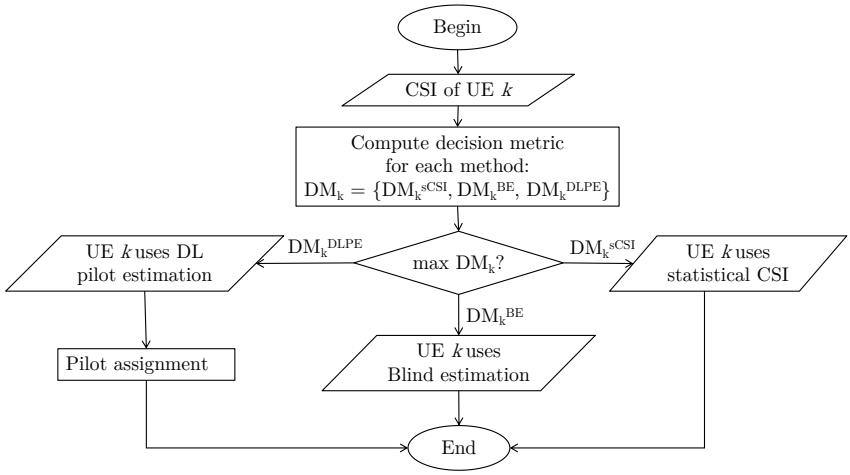


Figure 3.9: Proposed hybrid DL channel estimation strategy (Publication III).

In the proposed strategy, the decision metric can be computed based on CSI information at the CPU obtained after the UL training. As illustrated in Figure 3.9, the final decision involves selecting the estimation method that yields the highest decision metric value for each UE device. After all UE devices are assigned an estimation strategy, the CPU also performs DL pilot assignment for the UE devices designated to use DL pilot-based estimation. It is important to note that the number of DL pilots equals either the number of UE devices in the DL pilot group or a predefined maximum value. Then, DL pilot reuse is implemented when there are more UE devices than available pilots. Finally, the DL data and pilots are sent to the UE. These steps are summarized in Figure 3.10.

3.3.4 Results

The results presented in Publication III address different system configurations by varying the number of antennas per AP, the number of samples within the coherence interval, and the total number of UE devices. Along with the key performance metrics of NMSE and SE, the results also include the selection frequency. This metric evaluates the proportion of UE devices assigned to each estimation alternative across different Monte Carlo trials

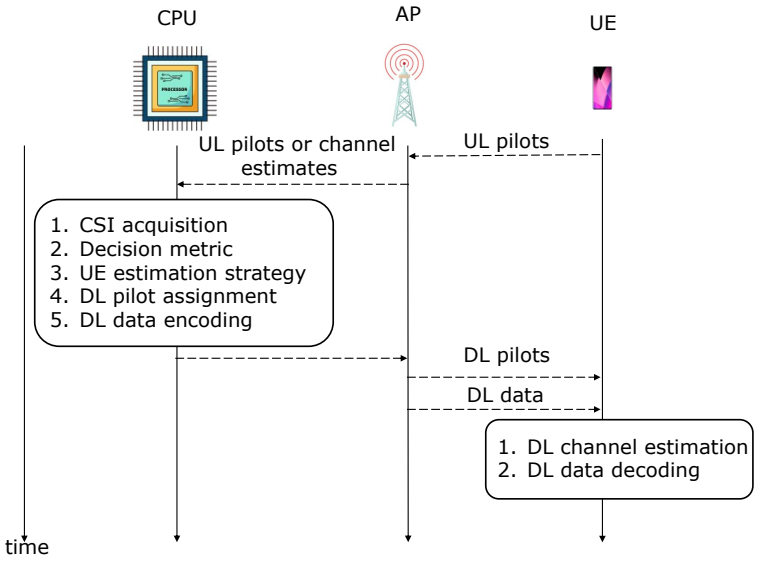
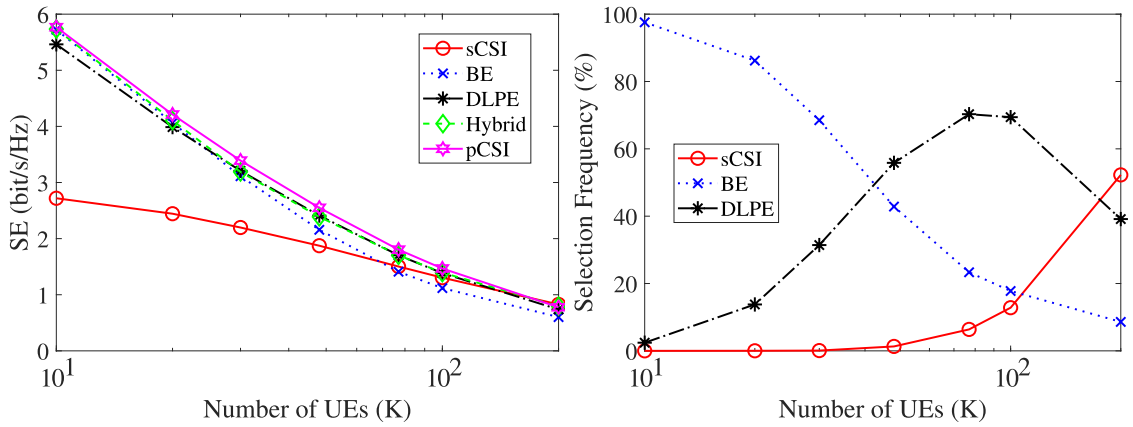


Figure 3.10: Communication diagram for the hybrid effective channel estimation.

and scenarios. In all scenarios, the proposed hybrid strategy yields the best SE performance compared with the other methods being used alone. Figure 3.11 depicts the case where the number of UE devices varies. Considering the average SE of the statistical CSI, blind estimation, and DL pilot-based estimation alternatives, it is worth noting that the best approach changes as the number of UE devices varies. On the other hand, the proposed hybrid channel estimation method can provide the best performance for any number of UE devices. This is reflected in the selection frequency result, where we can see that the proportion of UE devices using each alternative changes with the number of UE devices. Similar conclusions are also drawn for the other system configurations.



(a) Average SE (b) Frequency

Figure 3.11: Average SE and frequency of each estimation method being selected in the hybrid strategy by varying the number of UE devices (Publication III).

3.3.5 Conclusions

Publication III proposed a dynamic channel estimation method in UC CF massive MIMO networks that can change the channel estimation method of each UE according to the available resources and UE needs. The results demonstrated that the trade-off between accuracy and overhead is achieved to improve the overall data rate performance.

3.4 Publication IV: Trajectory Optimization in User-Centric Distributed Massive MIMO Systems Enabled by UAV Swarms

3.4.1 Research objectives

Publication IV explores the utilization of UAV swarms as the APs[1] in UC CF massive MIMO systems. The work discusses two scenarios: terrestrial network malfunction due to a disaster and congestion due to a UE surge. For these situations, UAV swarms can temporarily replace or assist the terrestrial infrastructure, as illustrated in Figure 3.12. The main objective of this research is to optimize the trajectory of UAVs to enhance SE and EE while accounting for the mobility of UE throughout the coverage area to emulate real-world situations.

Publication IV answers the following research question:

RQ 4: How can the positioning and trajectories of UAVs be optimized with a relatively low complexity in UC CF massive MIMO systems?

3.4.2 Rationale and context

There is a gap in the related technical literature when addressing trajectory optimization solutions in UC CF massive MIMO systems enabled by UAVs. The prevailing literature often assumes that the positions of UE devices are fixed and known by the optimization unit. This assumption neglects the mobility of UE devices when performing trajectory optimization. Publication IV addresses these limitations by proposing a heuristic trajectory optimization algorithm that does not require prior knowledge of the positions of UE devices and also works when considering UE mobility.

In this work, it is important to note that the UAVs are considered to be energy-limited. Additionally, the power for data transmission is usually on the milliwatt scale [15], while the required UAV propulsion power is in the watt range [38]. As a result, jointly optimizing power allocation for data transmission and UAV trajectory would not bring significant improvements to the performance of UC CF massive MIMO systems enabled by UAVs. Therefore, Publication IV also leverages consolidated scalable algorithms for power allocation, pilot assignment, and AP selection from the UC CF massive MIMO literature

[1] APs are referred to as transmission and reception points (TRPs) in Publication IV.

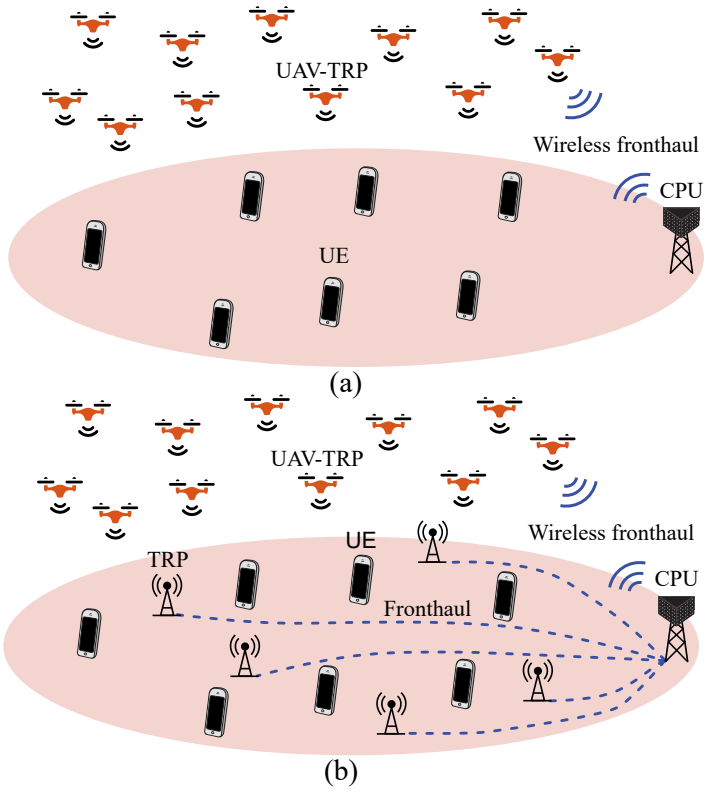


Figure 3.12: UC CF massive MIMO system enabled by a swarm of UAVs as aerial APs (Publication IV).

[15], [32]. This allows us to focus on the positioning and trajectory optimization of the UAVs.

3.4.3 Methodology

Publication IV introduces a UE mobility model based on a discrete-time Markov chain. In this model, the moving direction of each UE depends on the previous one, while the UE position is determined by its speed and current direction of movement. In addition, the work also leverages large-scale CSI to obtain insights into the SE performance metric at the CPU. To achieve this, closed-form expressions for SE are derived considering MR precoding, a Rician fading channel model, and the effect of fronthaul quantization. It is important to note that closed-form expressions are not available for non-MR precoding, but this performance metric is still useful to perform UAV positioning and trajectory optimization even when other precoding schemes are applied.

The optimization consists of a heuristic approach to optimize UAV positions, allowing for continuous iterative adjustments to UAV trajectories as UE devices move within the coverage area. The UAV trajectory optimization algorithm begins by selecting the subset of UAVs to adjust their positions within each episode. This subset is defined as UAVs

that serve the UE with the worst SE. In addition, an episode is the time frame where UE devices are assumed to remain at the same position, which is subdivided into iterations for position adjustment of each UAV. Specifically, the number of iterations is defined as the ratio between the UAV and UE velocities. Within an iteration, the selected UAV initially moves in a random direction, after which the SE is reevaluated. A new movement direction is computed using a truncated normal distribution, in which there is a higher probability of either maintaining the present direction or turning to the opposite direction. This decision depends on whether the previous action led to an improvement in SE or not. Note that only one UAV moves at each iteration in order to accurately assess the impact of this action. Figure 3.13 illustrates the proposed solution within each episode.

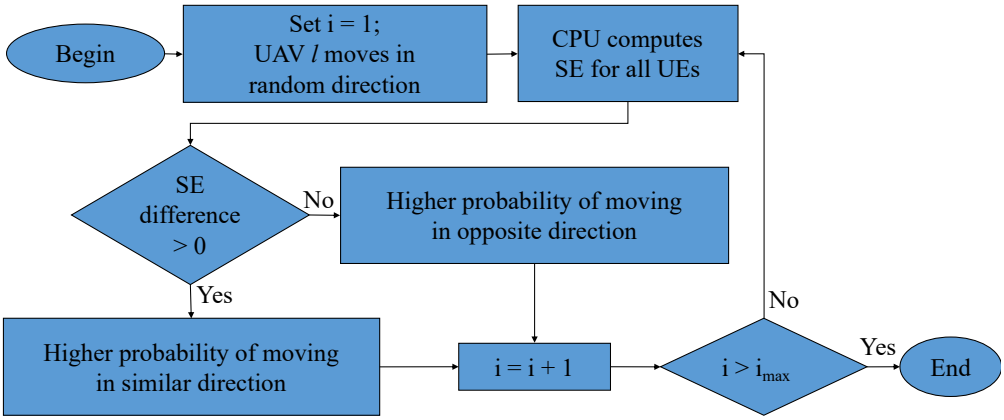


Figure 3.13: Iterative adjustment method for the positions of the selected UAVs per episode.

3.4.4 Results

The results in Publication IV emphasize the performance of the proposed solution in terms of SE, taking into account different precoding schemes. We also provide results for overall EE, power consumption, and the required fronthaul data rate. We evaluate different scenarios to account for varying UE densities: low, medium, and high. The number of serving APs is different for each UE density level. Specifically, it is computed based on preliminary simulations to obtain the expected average data rate performance of 200 Mbit/s per UE device. Moreover, we compare our solution with particle swarm optimization (PSO) and UAV fixed-position benchmark schemes. The PSO results assume that UE devices remain in the same position and that their positions are known in order to analytically compute the SE as the objective function. For that reason, the comparison with PSO exemplifies a UAV deployment scenario. This analysis demonstrates that the proposed solution performs well in comparison with PSO benchmark schemes and has the advantage that it does not require knowledge of the UE positions.

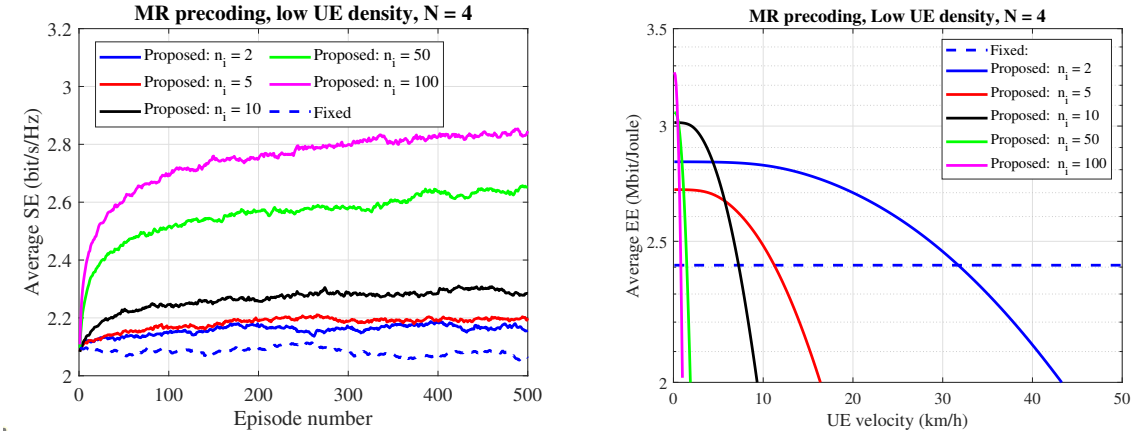
Additionally, we answer RQ 4 by evaluating the proposed UAV trajectory algorithm while the UE device moves across the coverage area. In this scenario, we compare the proposed

solution with the UAV fixed-position scheme. For this purpose, two case studies were conducted, and their main results are summarized below:

- UAVs are the only APs serving the UE;
- Both terrestrial and aerial APs serve the UE.

Case 1: UAVs are the only APs serving the UE

Figure 3.14 presents the SE and the EE for different numbers of iterations (n_i) in the proposed UAV trajectory optimization. In Figure 3.14a, it is observed that the performance of the proposed method improves with an increase in the number of iterations, because the algorithm has more flexibility to adjust the UAV positions. However, it is important to note that the number of iterations is not always a controllable design parameter. It is computed as the velocity ratio of UAVs and UE devices and represents the time frame where the UE positions do not change considerably. Thus, if the UE velocities are high, it may not be possible to deploy UAVs with even higher velocities. In addition, the required propulsion power of UAVs increases with their velocities, which may reduce the overall EE even though the SE improves. This behavior is demonstrated in Figure 3.14b by varying the maximum UE velocity. It can be noted that the EE of the proposed algorithm can be lower than that of the fixed positioning approach, depending on the UE velocities and the number of iterations n_i.



(a) (b)

Figure 3.14: (a) Average DL SE of the UE devices per episode for different numbers of iterations (n_i) in the proposed UAV trajectory optimization. (b) Average EE vs maximum velocity of the UE devices with varying numbers of iterations n_i (Publication IV).

Case 2: Both terrestrial and aerial APs serve the UE

Figure 3.15 shows the performance of the proposed method in the hybrid terrestrial and

aerial APs scenario for different precoding schemes. In a system composed only of terrestrial APs, these results illustrate how the UE performance in terms of SE decreases as the UE density increases from 130 to 200 UE devices. In this high UE density scenario, the UAVs are deployed to assist temporarily, and the results show that SE performance is restored when assuming MR precoding. For other precoding schemes, SE performances also increase, but not to the same level as with the medium UE density.

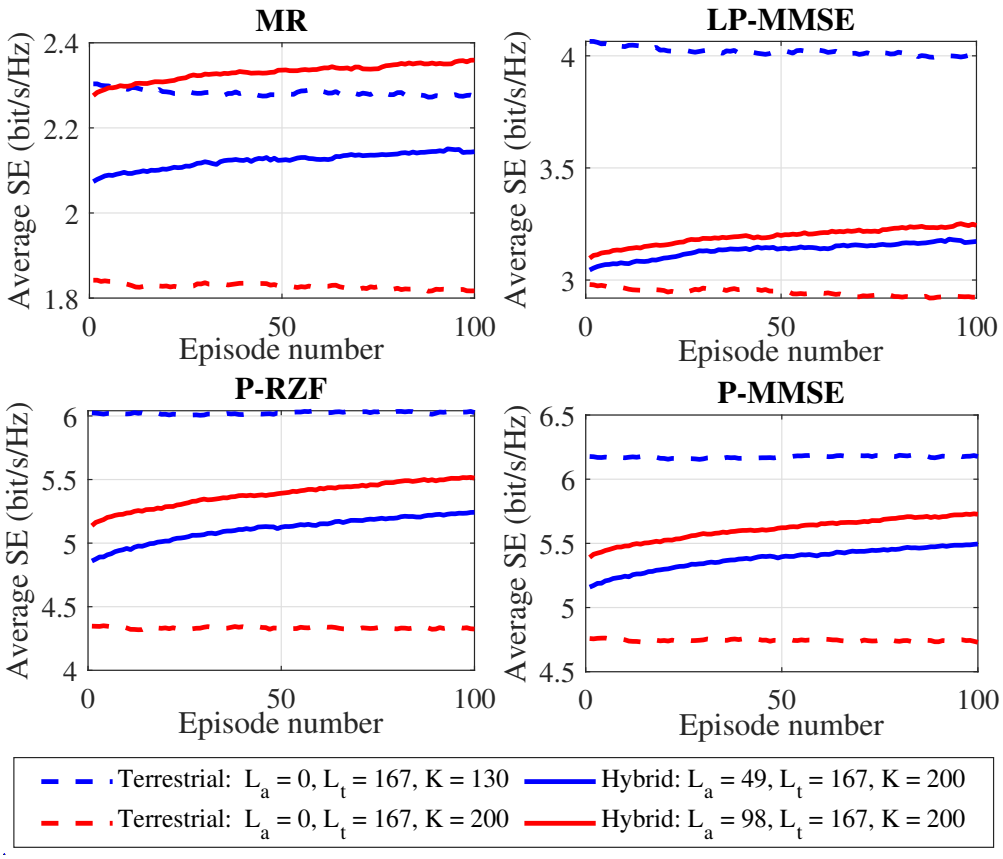


Figure 3.15: Average DL SE of the UE devices for systems consisting exclusively of terrestrial APs and those with a combination of terrestrial and aerial APs (Publication IV).

3.4.5 Conclusions

Publication IV showed that optimizing UAV trajectories in UC CF massive MIMO networks can lead to substantial improvements in SE. The findings suggested that integrating UAVs into communication networks can effectively address the challenges posed by UE mobility and UE demand. In summary, the SE result indicated that the proposed UAV trajectory optimization significantly improves SE, particularly in the scenarios of low UE density, and it performs better as the number of iterations increases. Additionally, the EE results demonstrated that UAVs should not move at high speeds to keep the propulsion

power low; thus, it is important to reduce the number of iterations when UE velocities increase.

3.5 Publication V: Scalable User-Centric Distributed Massive MIMO Systems With Restricted Processing Capacity

3.5.1 Research objectives

This work aims to evaluate the performance of scalable UC CF massive MIMO systems composed of multiple CPUs, as illustrated in Figure 3.16. Specifically, Publication V proposes a framework to ensure that the required CC and backhaul signaling are kept under control. Additionally, the work also proposes two solutions to adjust the AP[2] cluster according to the network implementations.

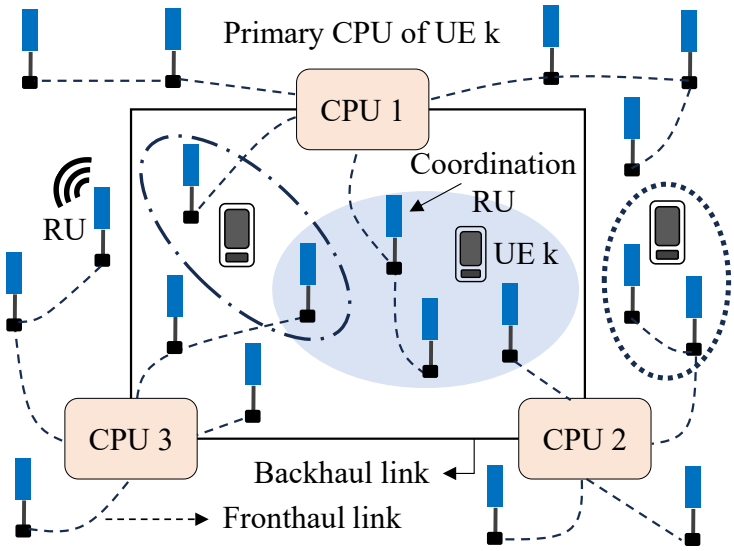


Figure 3.16: UC CF massive MIMO system with multiple CPUs. Each CPU is connected to a subset of APs, and the UE can be associated with APs linked to different CPUs (Publication V).

Publication V answers the following research question:

RQ 5: How can it be ensured that the computational complexity, signaling overhead, and energy consumption are kept under control in UC CF massive MIMO systems?

3.5.2 Rationale and context

UC CF massive MIMO systems, while offering enhanced SE and uniform service quality, often face challenges related to scalability due to the increasing resource demands

[2]APs are referred to as radio units (RUs) in Publication V.

as the network size grows. Although several approaches have been proposed to ensure scalability as the number of UE devices increases, there is a gap in the technical literature regarding the growing CC associated with signal processing as the number of APs increases. Additionally, the performance of the same AP cluster can vary significantly between different implementations. For instance, AP clusters with a large number of APs can negatively impact the EE and the CC of signal processing in the distributed implementation, while they can improve the SE for the centralized systems. Publication V addresses these challenges in UC systems composed of multiple CPUs.

The work also investigates how the multiple CPUs can coordinate to serve the UE and how the signaling can be handled in the backhaul links. Specifically, the CPUs may need to share signaling to serve the UE. This happens when the AP cluster of the UE consists of APs connected to different CPUs, as illustrated in Figure 3.16. This backhaul signaling is referred to as inter-CPU communication or inter-CPU coordination.

3.5.3 Methodology

Publication V first introduces the CPU requirements and backhaul signaling model for the scenario of multiple CPUs. As illustrated in Figure 3.16, one of the CPUs associated with the UE AP cluster is designated as its primary CPU and is responsible for the majority of its signal processing tasks. The remaining CPUs serving the UE are called secondary CPUs. The CPUs must coordinate and share their signal processing tasks efficiently to serve each UE device. In order to reduce inter-CPU communication, each CPU can only serve a limited number of UE devices as a secondary CPU, denoted as K^{sec}. Thus, the work proposes a method for the CPUs to adjust the AP cluster of each UE device to meet this condition.

Publication V proposes a framework that incorporates processing capacity and inter-CPU communication constraints. Two methods are proposed to limit the number of APs serving each UE device: one implemented at the CPUs and the other at the UE and APs. These methods aim to prevent the CC for channel estimation and signal precoding from increasing with the number of APs. For this purpose, we rely on a strategy where each UE device can be associated only with a finite number of APs, denoted as C_{max}. Figure 3.17 illustrates the maximum AP cluster size control procedure with CPU cooperation. In a nutshell, when a new UE device enters the network, it claims a master AP and performs any scalable UC AP selection scheme. The CPUs associated with the AP cluster of the UE compute the number of APs serving a given UE device. If the AP cluster size is less than the maximum, no action is required. Otherwise, the CPUs will drop the connection with the excessive APs. To drop the APs in excess, CPUs associated with each UE device sorts the channel gains in an ascending order and drops the connection of the first excessive APs presenting the smallest channel gains after the sort operation.

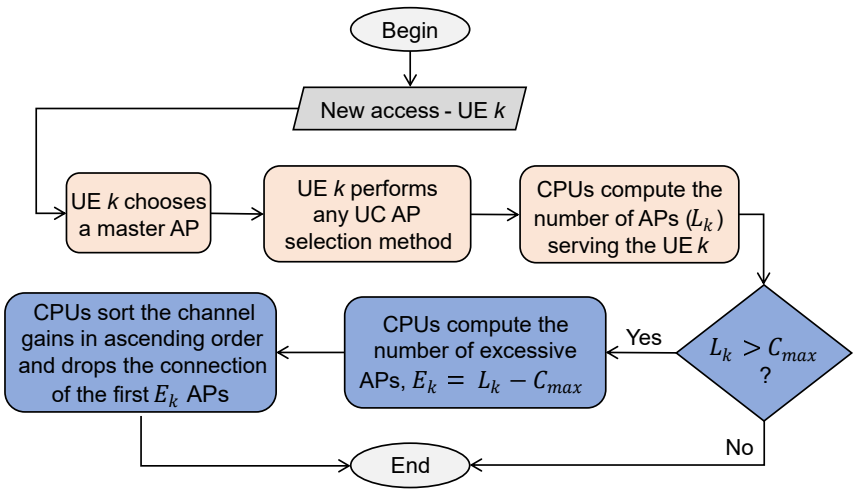


Figure 3.17: Flowchart of the AP cluster size control with CPU cooperation [3].

3.5.4 Results

The proposed approaches are compared with traditional scalable UC systems that do not perform AP cluster adjustments and do not limit processing demands. The effectiveness of the proposed solution is demonstrated through extensive numerical results, assuming distributed and centralized precoding schemes, and by simulating different use cases: varying the number of APs, UE devices, antennas per AP, and the maximum cluster size C_{max}. The results are illustrated in terms of SE, CC, backhaul traffic, and EE. Next, we discuss the main results that answer RQ 5.

Figure 3.18 presents the performance of the proposed approach (PA) to perform AP cluster size control with CPU cooperation, which is referred to as "with PA" in the results. In Figure 3.18a, the average SE shows that the proposed approach achieves similar performance to when there is no processing capacity limitation. The benefits of the proposed solution are evident in the CC to perform channel estimation and generate the combining vectors, which can be reduced by up to 98%, as shown in Figure 3.18b.

Moreover, it can be noted in Figure 3.18c that the proposed approach (i.e., processing capacity and inter-CPU communication restrictions) reduces the backhaul traffic significantly. Finally, Figure 3.18d illustrates the EE achieved by a UC system with and without the proposed approach for processing capacity limitations with different values of C_{max}. Note that the processing capacity limitation can provide a considerable improvement in the EE, especially for small values of C_{max}. This happens because the proposed strategy allows for reduced fronthaul and backhaul traffic, which, in turn, reduces the power consumption. In summary, the results in Figure 3.18 indicate that the proposed strategy allows UC CF massive MIMO systems to reduce their CC and signaling demands while keeping the SE under minor degradation and improving the EE.

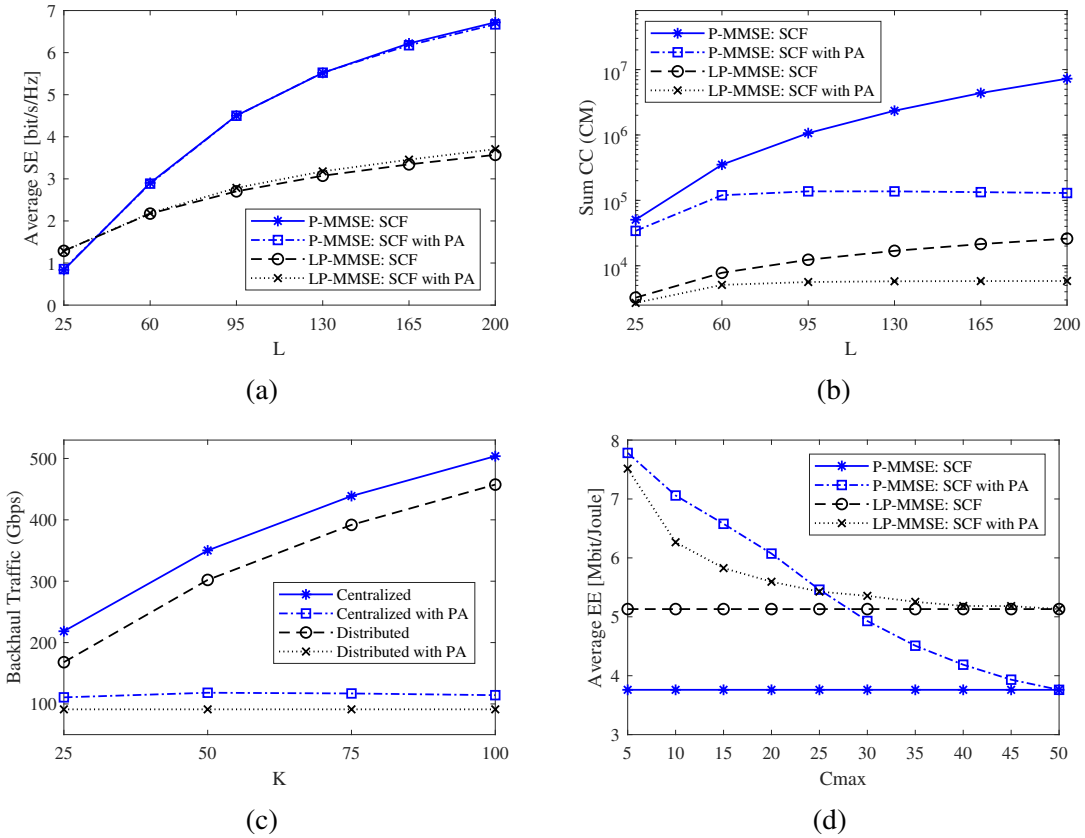


Figure 3.18: (a) Average SE and (b) CC in terms of the number of complex multiplications (CM) achieved by varying the number of APs L. (c) Backhaul traffic in each network implementation by varying the number of UE devices K. (d) Average EE achieved by varying C_{max} (Publication V).

3.5.5 Conclusions

Publication V demonstrated that scalable UC CF massive MIMO systems can achieve high performance levels while operating under limited processing capacities. By limiting the number of APs that each UE device connects to and implementing strategic AP cluster adjustments, the system can significantly reduce CC and enhance EE without substantial losses in SE. These findings provide valuable insights for designing scalable and efficient UC CF massive MIMO systems composed of multiple CPUs.

4 Conclusions and future work

This doctoral dissertation presented scalable solutions for enhancing UC CF massive MIMO systems. Specifically, the contributions of this work focused on effective channel estimation, trajectory optimization of UAVs as APs, and scalable AP selection strategies.

Publications I, II, III addressed DL effective channel estimation strategies. Publication I focused on a blind algorithm for effective channel estimation. Publication II proposed scalable pilot allocation and assignment methods for DL pilot-based estimation. Publication III aimed to adapt the choice of the estimation method for each UE device dynamically; i.e., a hybrid method that can be used either blind, DL pilot-based training, or statistical CSI methods. The numerical results demonstrated that blind estimation can outperform the statistical CSI solution even when the asymptotic arguments that inspire the algorithm do not hold. As for DL pilot-based estimation, it was demonstrated that it is the best method when the pilot overhead is small, because it generally has a higher estimation accuracy. The results of Publication II indicated that it is possible to achieve the best performance with any of the three estimation methods (statistical CSI, blind, and DL pilot-based) depending on the system parameters. This motivated the development of the proposed hybrid strategy in Publication III, and the results demonstrated that the solution can balance accuracy and overhead.

Publication IV addressed the trajectory optimization problem of UAVs jointly cooperating as the APs of a CF network. The work proposed a scalable UAV positioning and trajectory solution that can iteratively improve the SE performance of UE. Moreover, the proposed solution adjusts the UAV positions while the UE devices move across the coverage area. The results demonstrated that it is possible to improve the SE and EE of UC CF massive MIMO systems enabled by UAVs without requiring the knowledge of UE positions. However, the improvements are more significant in scenarios with low UE density and mobility.

Finally, Publication V aimed at adjusting AP cluster formation for reducing the CC and the signaling overhead in UC systems. The work analyzed systems composed of multiple CPUs and proposed two methods to limit the maximum number of APs serving each UE device. The results demonstrated the effectiveness of the proposed solution to reduce CC and improve EE. However, the AP cluster comprising a few APs can degrade the SE of a centralized implementation, although this impact is not significant in a distributed one.

4.1 Future work

To extend the contributions of this doctoral dissertation, some future research directions can be outlined.

Regarding the assumptions considered for the UC CF massive MIMO system model, fu-

ture work can also take into account imperfect knowledge of the channel statistics, hardware distortion, nonreciprocity, delay and latency, and channel aging due to UE mobility. These factors can influence the performance metrics such as SE and EE.

Regarding the DL effective channel estimation proposals, it is possible to modify the decision metric of the hybrid approach to explicitly include performance-related priorities, such as requirements for QoS and quality-of-experience (QoE). Additionally, future work can determine how to select the proper number of samples for the coherence interval based on UE mobility, taking into account the effect of channel aging. After that, the next task would be to investigate how to allocate the samples for channel estimation, both for UL and DL pilots.

It is also important to consider the computational and energy costs of the proposed algorithms for UAV trajectory and AP selection. For instance, computational complexity and control signaling requirements are factors that determine whether any resource allocation and management method or algorithm can be used in practice. However, this aspect is often neglected in the literature. In fact, the power required to apply resource allocation and management methods or algorithms is not considered for calculating energy efficiency, leading to its overestimation. Therefore, a future task would be to determine the feasibility of the proposed methods by comparing the computational and energy costs with the achieved benefits.

Additionally, future works can also conduct a techno-economic analysis of UC CF massive MIMO networks assisted by UAVs. This study requires estimating the total number of UAVs needed to operate within a given time frame, while accounting for possible replacements due to limited battery capacity, incorporating charging-rotation strategies, and exploring alternative energy solutions such as harvesting techniques.

References

[I] Souza, D. D., Freitas, M. M. M., Borges, G. S., Cavalcante, A. M., Costa, D. B. da, and Costa, J. C. W. A., Effective channel blind estimation in cell-free massive MIMO networks, *IEEE Wireless Commun. Lett.*, vol. 11, (3), pp. 468–472, Mar. 2022. DOI: 10.1109/LWC.2021.3132418.

[II] Souza, D. D., Freitas, M. M. M., Costa, D. B. da, *et al.*, 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. DOI: 10.1109/GLOBECOM48099.2022.10001150.

[III] Souza, D. D., Fernandes, A. L. P., Freitas, M. M. de, *et al.*, Effective channel hybrid estimation in User-Centric distributed massive MIMO networks, in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Mar. 2025, pp. 01–06. DOI: 10.1109/WCNC61545.2025.10978257.

[IV] Souza, D. D., Freitas, M. M. M., Fernandes, A. L. P., *et al.*, Trajectory optimization in user-centric distributed massive MIMO systems enabled by UAV swarms, *IEEE Transactions on Vehicular Technology*, vol. 74, (6), pp. 9252–9268, Jun. 2025. DOI: 10.1109/TVT.2025.3539120.

[V] Freitas, M. M. M., Souza, D. D., Fernandes, A. L. P., *et al.*, Scalable user-centric distributed massive MIMO systems with restricted processing capacity, *IEEE Trans. Wireless Commun.*, vol. 23, (12), pp. 19 933–19 949, Dec. 2024. DOI: 10.1109/TWC.2024.3491153.

[1] Freitas, M., Souza, D., Costa, D. B. da, *et al.*, Matched-decision AP selection for user-centric cell-free massive MIMO networks, *IEEE Trans. Veh. Technol.*, pp. 1–16, Jan. 2023. DOI: 10.1109/TVT.2023.3235980.

[2] Freitas, M. M. M., Souza, D. D., Costa, D. B. da, *et al.*, Reducing inter-CPU coordination in user-centric distributed massive MIMO networks, *IEEE Wireless Commun. Lett.*, pp. 1–5, 2023. DOI: 10.1109/LWC.2023.3252000.

[3] Freitas, M. M. M., Souza, D. D., Fernandes, A. L. P., *et al.*, Scalable user-centric distributed massive MIMO systems with limited processing capacity, in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 1–7.

[4] Fernandes, A. L. P., Souza, D. D., Costa, D. B. da, Cavalcante, A. M., and Costa, J. C. W. A., Cell-free massive MIMO with segmented fronthaul: Reliability and protection aspects, *IEEE Wireless Commun. Lett.*, pp. 1–1, 2022. DOI: 10.1109/LWC.2022.3166485.

[5] Fernandes, A. L. P., Souza, D. D., Natalino, C., *et al.*, A cost assessment methodology for user-centric distributed massive MIMO architectures, *IEEE Open j. Commun. Soc.*, vol. 5, pp. 3517–3543, 2024. DOI: 10.1109/OJCOMS.2024.3406374.

[6] Ueoka, T. T. R., Souza, D. D., Freitas, M. M., Cavalcante, A. M., and Costa, J. W., Performance of cell-free massive MIMO networks under Rayleigh and Rician fading, in *Anais do XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2022. DOI: 10.14209/sbrt.2022.1570811228.

[7] Ueoka, T. T. R., Costa, M. A. S., Souza, D. D., *et al.*, Pilot allocation and assignment optimization in user-centric distributed massive MIMO networks, in *Anais do XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2023. DOI: 10.14209/sbrt.2023.1570923555.

[8] Costa, M. A. S., Freitas, M. M., Ueoka, T. T. R., *et al.*, A fairer comparison of processing implementations in user-centric distributed massive MIMO systems, in *Anais do XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2023. DOI: 10.14209/sbrt.2023.1570915442.

[9] Interdonato, G., Björnson, E., Ngo, H. Q., Frenger, P. K., and Larsson, E. G., Ubiquitous cell-free massive MIMO communications, *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, p. 197, 2019. DOI: 10.1186/s13638-019-1507-0.

[10] Ngo, H. Q., Ashikhmin, A., Yang, H., Larsson, E. G., and Marzetta, T. L., Cell-free massive MIMO versus small cells, *IEEE Trans. Wireless Commun.*, vol. 16, (3), pp. 1834–1850, 2017. DOI: 10.1109/TWC.2017.2655515.

[11] Zhang, J., Chen, S., Lin, Y., Zheng, J., Ai, B., and Hanzo, L., Cell-free massive MIMO: A new next-generation paradigm, *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019. DOI: 10.1109/ACCESS.2019.2930208.

[12] Zhang, J., Björnson, E., Matthaiou, M., Ng, D. W. K., Yang, H., and Love, D. J., Prospective multiple antenna technologies for beyond 5G, *IEEE J. Sel. Areas Commun.*, vol. 38, (8), pp. 1637–1660, 2020. DOI: 10.1109/JSAC.2020.3000826.

[13] Björnson, E. and Sanguinetti, L., Scalable cell-free massive MIMO systems, *IEEE Trans. Commun.*, vol. 68, (7), pp. 4247–4261, 2020. DOI: 10.1109/TCOMM.2020.2987311.

[14] Björnson, E. and Sanguinetti, L., Making cell-free massive MIMO competitive with MMSE processing and centralized implementation, *IEEE Trans. Wireless Commun.*, vol. 19, (1), pp. 77–90, 2019.

[15] Demir, Ö., Björnson, E., and Sanguinetti, L., *Foundations of User-Centric Cell-Free Massive MIMO* (Foundations and Trends in Signal Processing). Hanover, MA, USA: Now Publ., 2021, ISBN: 9781680837902.

[16] Interdonato, G., Cell-free massive MIMO: Scalability, signal processing and power control, Ph.D. dissertation, Linköping University, Communication Systems, Faculty of Science & Engineering, 2020, p. 75, ISBN: 9789179298081. DOI: 10.3384/diss.diva-167218.

[17] Chen, Z. and Björnson, E., Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry, *IEEE Trans. Commun.*, vol. 66, (11), pp. 5205–5219, 2018. DOI: 10.1109/TCOMM.2018.2846272.

[18] Shin, W. and Vaezi, M., Uav-enabled cellular networks, in *5G and Beyond: Fundamentals and Standards*, Lin, X. and Lee, N., Eds. Cham: Springer International Publishing, 2021, pp. 165–200, ISBN: 978-3-030-58197-8. DOI: 10.1007/978-3-030-58197-8_6. [Online]. Available: https://doi.org/10.1007/978-3-030-58197-8_6.

[19] Valavanis, K. P. and Vachtsevanos, G. J., Eds., *Handbook of Unmanned Aerial Vehicles*. Springer Netherlands, 2015. DOI: 10.1007/978-90-481-9707-1. [Online]. Available: https://doi.org/10.1007/978-90-481-9707-1.

[20] Björnson, E., Hoydis, J., and Sanguinetti, L., *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*. Foundations and Trends in Signal Processing Series, Now Publishers, 2018, ISBN: 9781680833645.

[21] Jin, S.-N., Yue, D.-W., and Nguyen, H. H., Spectral and energy efficiency in cell-free massive MIMO systems over correlated rician fading, *IEEE Systems Journal*, vol. 15, (2), pp. 2822–2833, 2021. DOI: 10.1109/JSYST.2020.2993048.

[22] D'Andrea, C., Garcia-Rodriguez, A., Geraci, G., Giordano, L. G., and Buzzi, S., Cell-free massive MIMO for UAV communications, in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6. DOI: 10.1109/ICCW.2019.8756714.

[23] Polegre, A. A., Riera-Palou, F., Femenias, G., and Armada, A. G., 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, 2020. DOI: 10.1109/ACCESS.2020.3012736.

[24] Özdogan, Ö., Björnson, E., and Larsson, E. G., Massive MIMO with spatially correlated rician fading channels, *IEEE Trans. Commun.*, vol. 67, (5), pp. 3234–3250, 2019. DOI: 10.1109/TCOMM.2019.2893221.

[25] Björnson, E. and Jorswieck, E., *Optimal Resource Allocation in Coordinated Multi-cell Systems* (Foundations and Trends in Communications and Information). Hanover, MA, USA: Now Publ., 2013, ISBN: 9781601986382.

[26] Nayebi, E., Ashikhmin, A., Marzetta, T. L., Yang, H., and Rao, B. D., Precoding and power optimization in cell-free massive MIMO systems, *IEEE Trans. Wireless Commun.*, vol. 16, (7), pp. 4445–4459, Jul. 2017. DOI: 10.1109/TWC.2017.2698449.

[27] Ngo, H. Q. and Larsson, E. G., No downlink pilots are needed in TDD massive MIMO, *IEEE Trans. Wireless Commun.*, vol. 16, (5), pp. 2921–2935, 2017. DOI: 10.1109/TWC.2017.2672540.

[28] Interdonato, G., Ngo, H. Q., Frenger, P., and Larsson, E. G., Downlink training in cell-free massive MIMO: A blessing in disguise, *IEEE Trans. Wireless Commun.*, vol. 18, (11), pp. 5153–5169, 2019. DOI: 10.1109/TWC.2019.2933831.

[29] Björnson, E., Sanguinetti, L., Hoydis, J., and Debbah, M., Optimal design of energy-efficient multi-user MIMO systems is massive MIMO the answer? *IEEE Trans. Wireless Commun.*, vol. 14, (6), pp. 3059–3075, 2015. DOI: 10.1109/TWC.2015.2400437.

[30] Ngo, H. Q., Tran, L., Duong, T. Q., Matthaiou, M., and Larsson, E. G., On the total energy efficiency of cell-free massive MIMO, *IEEE Trans. Green Commun. Netw*, vol. 2, (1), pp. 25–39, 2018.

[31] Demir, Ö. T., Masoudi, M., Björnson, E., and Cavdar, C., Cell-free massive MIMO in O-RAN: Energy-aware joint orchestration of cloud, fronthaul, and radio resources, *IEEE J. Sel. Areas Commun.*, Jan. 2024. DOI: 10.1109/JSAC.2023.3336187.

[32] Interdonato, G., Frenger, P., and Larsson, E. G., Scalability aspects of cell-free massive MIMO, in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, Jul. 2019, pp. 1–6. DOI: 10.1109/ICC.2019.8761828.

[33] "Study on channel model for frequencies from 0.5 to 100 GHz, (release 16)," 3GPP, Sophia Antipolis, France, Tech. Rep. TR 38.901, 2019.

[34] Pasangi, P., Atashbar, M., and Mohassel Feghhi, M., Blind downlink channel estimation of multi-user multi-cell massive MIMO system in presence of the pilot contamination, *AEU - Int. J. Electron. Commun.*, vol. 117, p. 153 099, 2020, ISSN: 1434-8411. DOI: https://doi.org/10.1016/j.aeue.2020.153099.

[35] Ngo, H. Q., Larsson, E. G., and Marzetta, T. L., Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots, in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2013, pp. 293–298. DOI: 10.1109/Allerton.2013.6736537.

[36] Interdonato, G., Ngo, H. Q., Larsson, E. G., and Frenger, P., How much do downlink pilots improve cell-free massive MIMO? In *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–7. DOI: 10.1109/GLOCOM.2016.7841875.

[37] Interdonato, G., Frenger, P., and Larsson, E. G., Utility-based downlink pilot assignment in cell-free massive MIMO, in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, 2018, pp. 1–8.

[38] Zhang, J., Campbell, J. F., Sweeney II, D. C., and Hupman, A. C., Energy consumption models for delivery drones: A comparison and assessment, *Transp. Res. Part D: Transp. Environ.*, vol. 90, p. 102 668, Jan. 2021, ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2020.102668.

Publication I

Souza, D. D., Freitas, M. M. M., Borges, G. S., Cavalcante, A. M., Costa, D. B. da, and Costa, J. C. W. A.
Effective Channel Blind Estimation in Cell-Free Massive MIMO Networks

Reprinted with permission from
IEEE Wireless Communications Letters
Vol. 11, no. 3, pp. 468-472, March 2022

Effective Channel Blind Estimation in Cell-Free Massive MIMO Networks

Daynara D. Souza, Marx M. M. Freitas, Gilvan S. Borges, André M. Cavalcante, *Member, IEEE*, Daniel B. da Costa, *Senior Member, IEEE*, and João C. W. A. Costa, *Senior Member, IEEE*

***Abstract*—This letter investigates the performance of a new blind method for effective channel estimation in user-centric cell-free massive multiple-input multiple-output (MIMO) networks. To this end, comparisons with the perfect and statistical channel state information (CSI) methods are performed assuming different precoding choices. The results demonstrate that blind estimation can significantly reduce normalized mean-square error and improve spectral efficiency compared with the statistical CSI method. Additionally, blind estimation works even when the asymptotic arguments that inspire the algorithm do not hold, performs reasonably well in the presence of pilot contamination, and can boost the performance of systems using maximum-ratio (MR) precoding.**

***Index Terms*—Blind estimation, cell-free massive MIMO networks, effective channel estimation, statistical CSI.**

I. Introduction

USER-CENTRIC (UC) cell-free massive multiple-input multiple-output (MIMO) systems consist of geographically distributed access points (APs) that cooperate to serve each user equipment (UE). In UC networks, each UE is served by a subset of APs, which reduces the fronthaul signaling for channel state information (CSI) and data sharing. Moreover, they can provide higher macro-diversity and higher coverage probability than co-located massive MIMO systems, where all the antennas are placed in a compact area. Consequently, UC systems potentially improve the performance of the worst UEs that would be located at cell-edges [1], [2].

Considering time-division duplex (TDD), the system can precode the downlink (DL) data based on uplink (UL) channel estimates, relying on channel reciprocity property. In this way, the data symbol intended for a given UE is scaled only by the *effective channel*. The UEs need to know their effective channels to decode the corresponding DL data symbols coherently [3]. Conventionally, the UEs do not need to estimate the effective channel in massive MIMO systems, as the channel hardening feature enables data decoding based on statistical CSI. Essentially, channel hardening is a phenomenon that makes a random channel behaves almost deterministically when the number of antennas is large [3], [4]. However, it is less pronounced in UC systems due to the geographical distribution of the APs as well as by the fact that the UEs connect only to a subset of APs equipped with few antennas [2]–[4]. Accordingly, the UEs may need to estimate the effective channel in UC systems. To this end, two common approaches employed in the literature refer to the use of DL pilots and the proposal of blind algorithms. A drawback of the DL pilots-based approach is that it has to use additional resource samples on training pilots. On the other hand, blind estimation does not require any samples, it uses the average power of the received DL signals to estimate the effective channel. This idea was proposed for cellular massive MIMO and investigated in [5], [6].

This letter investigates the performance of the blind estimation approach in UC cell-free massive MIMO networks. To the best of the authors' knowledge, such analysis has not been carried out in the literature yet. Hence, the main contributions of this letter can be summarized as:

- It is proposed a blind estimation approach for a UC cell-free massive MIMO network that supports arbitrary channel types, precoding technique, pilot reuse method, and AP selection methods.
- Numerical results showing the effectiveness of the blind estimation method in terms of normalized mean square error (NMSE) and achievable DL spectral efficiency (SE) are provided. The impact of different channel hardening conditions and pilot contamination is investigated by varying the number of antennas per AP and increasing pilot reuse, respectively. The results are shown by considering different precoding methods, namely, maximum ratio (MR) and local partial MMSE (LP-MMSE).

Notation: Boldface lowercase and uppercase letters denote column vectors and matrices, respectively. The superscripts $()^{\mathrm{T}}$ and $()^{\mathrm{H}}$ denote transpose and conjugate-transpose, respectively. The trace, Euclidean norm and expectation operator are denoted as $\mathrm{tr}(\,.\,)$, $\|\,.\,\|$ and $\mathbb{E}\{\,.\,\}$, respectively. The notation $\xrightarrow{P}$ means convergence in probability and $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ is a complex Gaussian random variable with μ mean and σ^2 variance.

II. System Model

We assume a TDD cell-free massive MIMO system consisting of L APs, equipped with N antennas each, and K

Manuscript received November 19, 2021; accepted November 30, 2021. Date of publication December 3, 2021; date of current version March 9, 2022. This work was supported in part by the Innovation Center, Ericsson Telecomunicações S.A, Brazil; in part by the National Council for Scientific and Technological Development (CNPq); and in part by the Coordination for the Improvement of Higher Education Personnel (CAPES). The associate editor coordinating the review of this article and approving it for publication was G. Alexandropoulos. *(Corresponding author: Daniel B. da Costa.)*

Daynara D. Souza, Marx M. M. Freitas, and João C. W. A. Costa are with the Applied Electromagnetism Laboratory, Federal University of Pará—UFPA, Belém 66075-110, Brazil (e-mail: daynara@ufpa.br; marx@ufpa.br; jweyl@ufpa.br).

Gilvan S. Borges is with the Department of Education, Control and Industrial Processes, Information and Communication, Federal Institute of Pará—IFPA, Belém 66093-020, Brazil (e-mail: gilvan.borges@ifpa.edu.br).

André M. Cavalcante is with the Ericsson Research, Ericsson Telecomunicações S.A., Indaiatuba 13337-300, Brazil (e-mail: andre.mendes.cavalcante@ericsson.com).

Daniel B. da Costa is with the Future Technology Research Center, National Yunlin University of Science and Technology, Douliu 64002, Taiwan, and also with the Department of Computer Engineering, Federal University of Ceará, Sobral 62010-560, Brazil (e-mail: danielbcosta@ieee.org).

Digital Object Identifier 10.1109/LWC.2021.3132418

single-antenna UEs. The total number of antennas in the network is $M = NL$, where $M > K$. The APs are connected to a central processing unit (CPU) through fronthaul links that are assumed to be error-free and able to support the data traffic [7]. We consider a TDD frame whose length equals the coherence interval, which means that the channel is static within a frame but varies independently from frame to frame. The channel vector $\mathbf{h}_{kl} \in \mathbb{C}^{N\times 1}$ between the AP l and UE k (and vice-versa) undergoes independent correlated Rayleigh fading, being defined as $\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{kl})$, where $\mathbf{R}_{kl} \in \mathbb{C}^{N\times N}$ is the statistical spatial correlation matrix, which models the large-scale fading, considering spatial channel correlation, path loss, and shadowing [2], [7].

Each coherence block comprises τ_c samples, where τ_{up} samples are reserved for UL pilot signals and τ_d for DL data transmissions. During the UL training phase, all UEs simultaneously send pilot sequences of τ_{up}-length to the APs. Then, the UL channels are estimated by correlating the received signals with a known pilot sequence and performing minimum mean square error (MMSE) estimation. The pilot signals are pair-wisely orthogonal, and different UEs can be assigned to the same pilot sequence, when $K > \tau_{up}$. Let $t_k \in \{1, \ldots, \tau_{up}\}$ denote the index of the pilot assigned to UE k, the subset of UEs that uses the same pilot as UE k is denoted as $\mathcal{P}_k \subset \{1, \ldots, K\}$, including itself. The correlated received pilot signal at AP l can be expressed as [2]

$$\mathbf{y}^{\text{pilot}}_{t_k l} = \sum_{i\in\mathcal{P}_k} \sqrt{\tau_{up}\eta_i}\, \mathbf{h}_{il} + \mathbf{n}_{t_k l}, \tag{1}$$

where η_i represents the UL transmit power of UE i, and $\mathbf{n}_{t_k l} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma^2_{ul}\mathbf{I}_N)$ stands for the noise term. The MMSE channel estimate of $\mathbf{h}_{kl}$ based on $\mathbf{y}^{\text{pilot}}_{t_k l}$ is given by

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_{up}\eta_k}\mathbf{R}_{kl}\boldsymbol{\Psi}^{-1}_{t_k l}\mathbf{y}^{\text{pilot}}_{t_k l}, \tag{2}$$

where $\boldsymbol{\Psi}_{t_k l} = \sum_{i\in\mathcal{P}_k} \eta_i \tau_{up}\mathbf{R}_{il} + \sigma^2_{\text{ul}}\mathbf{I}_N$ is the correlation matrix of the received signal in (1).

In the DL mode, each AP serves the UEs implementing power control and precoding based on the UL estimates. Recall that in UC systems only a subset of APs serves the UE k, which is denoted as $\mathcal{M}_k \subset \{1, \ldots, L\}$. A diagonal matrix $\mathbf{D}_{kl} \in \mathbb{N}^{N\times N}$ is employed to represent which APs are serving the UE k, being defined as

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & l \in \mathcal{M}_k \\ \mathbf{0}_N & l \notin \mathcal{M}_k. \end{cases} \tag{3}$$

Let $q_k(n)$ be the n-th symbol intended for UE k. We assume that $\mathbb{E}\{\mathbf{q}(n)\mathbf{q}(n)^H\} = \mathbf{I}_K$, where $\mathbf{q}(n) \triangleq [q_1(n), \ldots, q_K(n)]^T$. Thus, the data signal sent by AP l can be written as

$$\mathbf{x}_l(n) = \sum_{k=1}^{K} \mathbf{D}_{kl}\mathbf{w}_{kl}\, q_k(n), \tag{4}$$

where $\mathbf{w}_{kl} \in \mathbb{C}^{N\times 1}$ represents the precoding vector, such that $\mathbb{E}\{\|\mathbf{w}_{kl}\|^2\} = \rho_{kl}$, with ρ_{kl} being the transmit power that AP l assigns to UE k. After the signal propagates through the channel, UE k receives a linear combination of the signals transmitted by the APs, i.e.,

$$\begin{aligned} y_{d,k}(n) &= \sum_{l=1}^{L} \mathbf{h}_{kl}\mathbf{x}_l(n) + n_{d,k}(n) \\ &= \underbrace{\alpha_{kk}\, q_k(n)}_{\text{desired signal}} + \underbrace{\sum_{k'\neq k}^{K} \alpha_{kk'}\, q_{k'}(n)}_{\text{inter-user interference}} + \underbrace{n_{d,k}(n)}_{\text{noise}}, \end{aligned} \tag{5}$$

where

$$\alpha_{kk'} = \sum_{l=1}^{L} \mathbf{h}^{\text{H}}_{kl}\mathbf{D}_{k'l}\mathbf{w}_{k'l}, \quad k' = 1, \ldots, K. \tag{6}$$

The noise at the receiver is formulated as $n_{d,k}(n) \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2_{dl})$. In addition, we call α_{kk} the effective channel for UE k, and $\alpha_{kk'}, k' \neq k$, the effective interfering channel.

III. Blind Channel Estimation

A. Proposed Blind Channel Estimation

In order to coherently detect the transmitted data symbol q_k, UE k needs to know the effective channel α_{kk}, which can be acquired by estimation methods or relying on statistical CSI [2], [3], [5]–[7]. Herein, we will propose a blind estimation of the effective channel gain α_{kk} in order to not require any DL pilots. The first step of the blind estimation algorithm is to compute the sample average power of the received signal at UE k per coherence interval, i.e.,

$$\xi_k = \frac{\sum_{n=1}^{\tau_d} |y_{d,k}(n)|^2}{\tau_d}, \tag{7}$$

where $y_{d,k}(n)$ denotes the n-th sample received at UE k, shown in (5), and $\tau_d = \tau_c - \tau_{up}$ is the number of symbols per coherence interval used for DL data transmission. It is expected that the channel and noise randomness average out in ξ_k when the length of the coherence interval is large. Therefore, from the law of large numbers, for $\tau_d \to \infty$, we have

$$\xi_k - \quad |\alpha_{kk}|^2 + \sum_{k'\neq k}^{K} |\alpha_{kk'}|^2 + \sigma^2_{dl} \quad \xrightarrow{P} 0. \tag{8}$$

There are many unknowns in Eq. (8) that prevent computing α_{kk} from ξ_k. To solve this issue, one can use once again the law of large numbers to approximate the inter-user interference term $\sum_{k'\neq k}^{K} |\alpha_{kk'}|^2$ by its mean [5]. As a consequence, when K and τ_d are large, ξ_k in (7) can be approximated by

$$\xi_k \approx \quad |\alpha_{kk}|^2 + \sum_{k'\neq k}^{K} \mathbb{E}\Big\{|\alpha_{kk'}|^2\Big\} + \sigma^2_{dl} \quad . \tag{9}$$

The probability that the real part of α_{kk} is much larger than the imaginary part of α_{kk} increases when the number of APs serving each UE end/or antennas per AP are large, i.e., as the channel hardening improves [2], [3], [5]. Based on this, the effective channel in (9) can be approximated by

$$\alpha_{kk} \approx |\alpha_{kk}| \approx \sqrt{\xi_k - Z_k}, \tag{10}$$

where

$$Z_k = \sum_{k' \neq k}^{K} \mathbb{E}\Big\{ |\alpha_{kk'}|^2 \Big\} + \sigma_{dl}^2. \tag{11}$$

Considering the case where the square root argument in (10) is non-positive, the statistical CSI approach that estimates the effective channel by its mean value is used, and the effective channel blind estimator can finally be formulated by

$$\hat{\alpha}_{kk} = \begin{cases} \sqrt{\xi_k - Z_k} & \text{if } \xi_k > Z_k \\ \mathbb{E}\{\alpha_{kk}\}, & \text{otherwise.} \end{cases} \tag{12}$$

Such estimator is the same as [5], but for a UC cell-free massive MIMO system instead of the co-located one. We assume that the UE knows $\mathbb{E}\{\alpha_{kk}\}$ and Z_k since they are functions only of the large-scale coefficients, which are constant over many coherence intervals [2], [5].

B. Normalized Mean Squarer Error and Spectral Efficiency

We compute the NMSE between the channel estimate $\hat{\alpha}_{kk}$ and the effective channel α_{kk} as

$$\mathrm{NMSE}_k = \frac{\mathbb{E}\Big\{ |\alpha_{kk} - \hat{\alpha}_{kk}|^2 \Big\}}{\mathbb{E}\Big\{ |\alpha_{kk}|^2 \Big\}}. \tag{13}$$

Before deriving the achievable DL SE, note that $\hat{\alpha}_{kk}$ is correlated with the data signal, which makes too intractable the derivation of SE. As in [5], we rewrite ξ_k in (7) to

$$\xi_k'(n) = \frac{\sum_{n'=1, n' \neq n}^{\tau_d} \big| y_{d,k}(n') \big|^2}{\tau_d - 1}. \tag{14}$$

Thus, the estimate $\hat{\hat{\alpha}}_{kk}$ is computed using $\xi_k'(n)$, which is close to $\hat{\alpha}_{kk}$ when τ_d is large. The achievable DL SE for UE k based on the use-and-then-forget (UatF) lower bound can be computed as [3], [5]

$$\mathrm{SE}_k = \left(1 - \frac{\tau_{up}}{\tau_c} \right) \log_2(1 + \mathrm{SINR}_k), \tag{15}$$

where

$$\mathrm{SINR}_k = \frac{|\mathbb{E}\{\frac{\alpha_{kk}}{\hat{\alpha}_{kk}}\}|^2}{\mathrm{Var}\Big\{ \frac{\alpha_{kk}}{\hat{\alpha}_{kk}} \Big\} + \sum\limits_{k' \neq k}^{K} \mathbb{E}\Big\{ |\frac{\alpha_{kk'}}{\hat{\alpha}_{kk}}|^2 \Big\} + \sigma_{dl}^2 \mathbb{E}\Big\{ \frac{1}{|\hat{\alpha}_{kk}|^2} \Big\}} \tag{16}$$

is the signal-to-interference-plus-noise ratio (SINR) of UE k. Other approaches could be used to compute the achievable SE, e.g., the methods in [3, Sec. III-A], and [5, Sec. V], but the first one is not a rigorous lower bound while the second one requires complicated numerical computations, unlike the UatF method used in this work.

IV. Numerical Results

This section presents illustrative numerical examples to show the efficiency of the proposed blind estimation method in UC cell-free massive MIMO networks. Unless stated otherwise, the main simulation parameters are $L = 100$ APs, with each one being equipped with $N = 4$ antennas, and $K = 20$ UEs. It is considered $\tau_c = 200$, $\tau_{up} = 10$, and $\tau_d = 190$.

TABLE I
Parameters Assumed for the UMi Path Loss and Local Scattering Spatial Correlation Model

Parameter	Value
Shadow fading standard deviation, σ_{SF}	4 dB
AP/UE antenna height, $h_{\mathrm{AP}}, h_{\mathrm{UE}}$	11.65 m, 1.65 m
RX noise figure (NF)	8 dB
Carrier frequency, bandwidth	3.5GHz, 100MHz
Angular standard deviations (ASDs)	$\sigma_\varphi = \sigma_\theta = 15^\circ$
Antenna spacing	1/2 wavelength distance

Since the analytical steps of the blind method are inspired by asymptotic arguments, the number of DL data samples τ_d, UEs K and antennas per AP N are varied throughout the results. The maximum transmission powers of the UEs and APs are 100 mW and 200 mW, respectively. We assume that the UEs are uniformly distributed into the square coverage area of 1 km $\times$ 1 km and the APs are placed following a hard core point process (HCPP).[1] The results are obtained through Monte-Carlo simulations to account for different AP/UE locations and channel realizations. These simulation parameters are in accordance with [2]–[4].

We used a joint pilot assignment and AP selection scheme presented in [2, Sec. 4.4]. In this method, the first τ_{up} UEs are assigned mutually orthogonal pilots, and the remaining UEs are assigned to the pilot that causes the lowest pilot contamination. After that, each AP chooses to serve τ_{up} UEs with the largest channel gain in each pilot. This method takes into account both the pilot contamination and scalability issues since the complexity and resource requirements for each AP remain finite when the number of UEs goes to infinity. Although it is not optimal, the SE performance loss compared to the case where all UEs are served by all APs is small [2]. It is considered a distributed network implementation, where each AP performs the channel estimation, precoding and power allocation locally. In order to mitigate interference, it is used MR and LP-MMSE precoding schemes, and for power allocation, we use the heuristic method presented in [2, Sec. 6.3], which allocates power directly proportional to the large-scale fading gains. The shadowing terms of an AP to different UEs are correlated and the correlation matrices $\mathbf{R}_{mk}$ follow the local scattering spatial correlation model, both computed as in [2]. The propagation model adopted is the 3GPP Urban Micro (UMi) path loss model with line-of-sight (LOS)/non-line-of-sight (NLOS) conditions defined in 3GPP TR 38.901 [8]. Table I specifies the parameters used as entries for the UMi and $\mathbf{R}_{mk}$ models [2], [9].

Fig. 1 depicts the cumulative distribution function (CDF) of the NMSE between the effective channel and its estimate per UE. One can observe that the blind estimation scheme achieves the best NMSE performance for both MR and LP-MMSE, even for small values of τ_d. The curve for $\tau_d = \infty$ is used to evaluate the blind estimator when the coherence interval is

[1] In this method, the APs locations are required to have a spacing regularity, which means that the distance between any two APs cannot be smaller than $d_{\min} = \sqrt{A/L}$, 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 rate $1/d_{\min}$, then randomly update the location of APs that do not meet the spacing requirement until it is fulfilled.

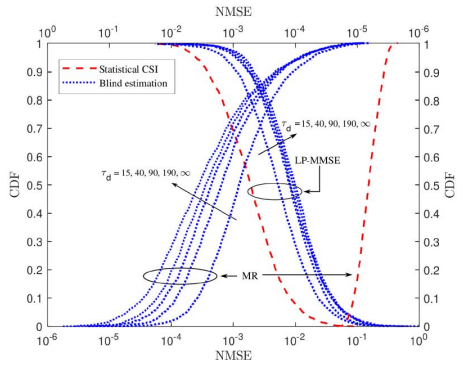


Fig. 1. CDF of the NMSE between the effective channel and its estimate per UE.

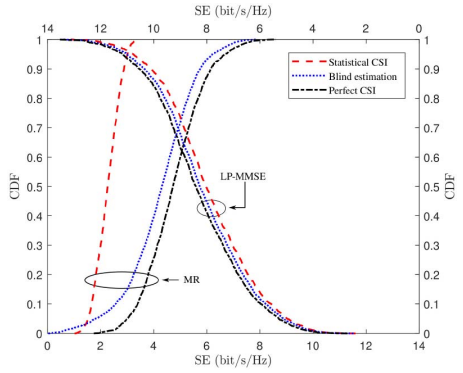


Fig. 2. CDF of the achievable DL SE per UE.

infinitely large, which means that the UE knows the asymptotic value of ξ_k. For MR precoding, the gap between blind estimation with $\tau_d = 190$ and its asymptotic result shows that slightly better results can be achieved if τ_d is larger. On the other hand, the blind estimation with $\tau_d = 190$ almost matches its asymptotic result for LP-MMSE.

In Fig. 2, we compare the CDF of the achievable DL SE per UE of blind estimation with two methods, statistical CSI and perfect CSI. The conventional statistical CSI method uses average effective channel $\mathbb{E}\{\alpha_{kk}\}$ as the estimates, and an achievable DL SE can be computed using the *hardening bound* [1], [2]. The perfect CSI curves represent the SE when the UE has perfect knowledge of the effective channel, achieved in a genie-aided manner [2]. For MR precoding, blind estimation improves the system performance significantly compared with statistical CSI, while the improvement is smaller for LP-MMSE. For instance, with MR precoding blind estimation can increase about 91.57% the SE of the 50% likely UEs, and higher gains are achieved for UEs with better channel conditions. On the other hand, the gain is around 3% for LP-MMSE. It is expected that the need to estimate the effective channel with LP-MMSE becomes lower because this robust precoding reduces the effective channel variations, which means it has a higher degree of channel hardening than MR [1]. This is in accordance with [5], where blind estimation provides higher gains in lower channel hardening conditions.

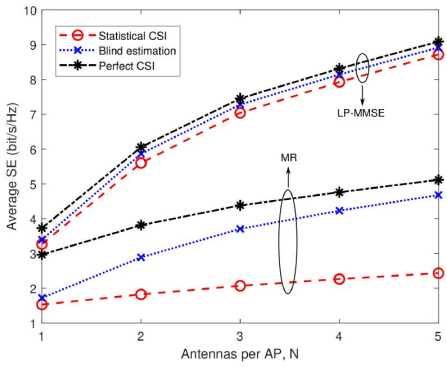


Fig. 3. Average SE versus the number of antennas per AP, *N*.

For MR, one can also observe that blind estimation provides lower SE than statistical CSI for the worst UEs, even though the NMSE is better. The reason is that the UatF bound can underestimate the SE in scenarios with low channel hardening [5]. The main reason for using the UatF method is that it is a consolidated and straightforward way to obtain a rigorous lower bound on the capacity, which means that it always provides results that are lower than or equal to the perfect CSI curves.

Fig. 3 exhibits the average SE versus the number of antennas per AP. Increasing N improves the channel hardening and the assumption that the real part of the effective channel is larger than the imaginary becomes true. As expected, the increase of N potentiates the SE in all estimation schemes. For MR, the blind estimation method has the most significant increase in SE from $N = 1$ to $N = 5$. This occurs because the UatF capacity bound becomes tighter and the blind estimation scheme more accurate as N increases. These facts lead to a higher and more accurate SE. On the other hand, the gains afforded by the blind estimation scheme are not expressive for LP-MMSE. The increased channel hardening provided by LP-MMSE allows systems using statistical CSI and blind estimation to provide similar SE performance. Therefore, one can note that the performance of the blind estimation algorithm in UC cell-free massive MIMO is strongly affected by the precoding choice, and it is most recommended for less robust schemes like MR.

In Fig. 4, the average NMSE between the effective channel and its estimate is plotted as a function of the number of UEs K. For LP-MMSE, increasing K makes the approximation of $\sum_{k' \neq k}^{K} |\alpha_{kk'}|^2$ by its mean better, this is why the NMSE initially decreases as K increases. However, pilot contamination occurs when $K > 10$ since the number of pilots is fixed at $\tau_{up} = 10$, which worsens this approximation, consequently increasing the NMSE. Since MR precoding is worse at suppressing interference than LP-MMSE, the interference caused by having more UEs overcomes the benefit of the better approximation and degrades the NMSE. For both precoders, the results indicate that the inter-user interference term $\sum_{k' \neq k}^{K} |\alpha_{kk'}|^2$ can be approximated by its mean even for small K, since blind estimation provides a lower average NMSE than statistical CSI for almost all considered values.

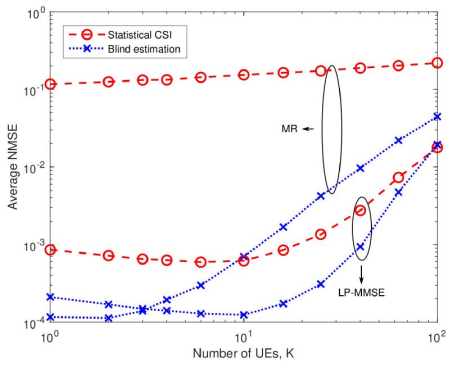


Fig. 4. Average NMSE between the effective channel and its estimate versus the number of UEs, K.

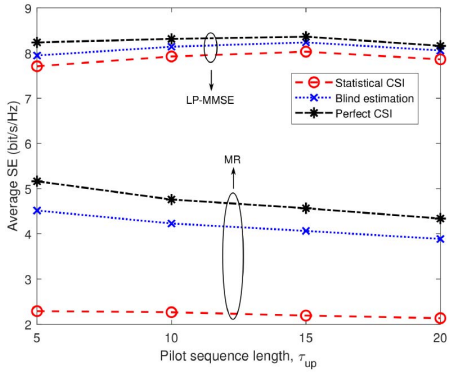


Fig. 5. Average SE versus the pilot sequence length, τ_{up}.

The only exception is when $K = 100$ using LP-MMSE, as the average NMSE is almost the same for both estimation methods. However, additional results revealed that the NMSE for 90% of UEs with blind estimation is still smaller than with statistical CSI, but the NMSEs of the other 10% of UEs are significantly higher. This indicates the need to better select the switching point between the two options in Eq. (12) to avoid that. The inter-user interference and its mean are zero when $K = 1$, but the NMSE is not zero because the blind method can only estimate the absolute value of the effective channel. The NMSE for statistical CSI is also not zero, as it depends on the quality of UL channel estimation and the channel hardening condition.

Fig. 5 shows the average SE versus the pilot sequence length. Since we consider $K = 20$, the pilots are reused among the UEs if $K > \tau_{up}$. As well-known, pilot reuse causes coherent interference and decreases the UL channel estimation quality, which degrades the achievable DL SE [2], [7]. For MR precoding, decreasing pilot reuse (increasing τ_{up}) does not improve the average SE since the estimation overhead decreases the pre-log factor in (15), even though there is less coherent interference and better channel estimation. On the other hand, the best pilot sequence length is a middle ground between overhead and estimation quality for LP-MMSE. This is because the LP-MMSE is more sensitive to the UL channel estimation quality, since its ability to mitigate interference depends on how good channel estimates are. One can note that the SE achieved by the blind estimation is not significantly affected by UL pilot contamination, even though the NMSE increases. The UL pilot contamination affects the SE of all methods practically in the same way. Pilot contamination does not significantly impact the SE because the used pilot assignment strategy minimizes pilot contamination interference.

V. Conclusion

This letter proposed a blind method for effective channel estimation in UC cell-free massive MIMO systems. The blind algorithm was compared with the conventional statistical CSI method and the perfect CSI scheme. We analyzed the results in terms of NMSE and achievable DL SE for MR and LP-MMSE precoding techniques, and varied the numbers of DL data samples τ_d, UEs K, antennas per AP N, and pilot sequence length τ_{up}. The results demonstrated that blind estimation can reduce the NMSE and increase about 91.57% the SE of the 50% likely UEs compared with the statistical CSI approach. Blind estimation can outperform the statistical CSI solution even when the asymptotic arguments that inspire the algorithm do not hold, i.e., even when τ_d, K and N are small. Moreover, it improves the SE of systems that employ MR precoding technique and performs reasonably well in the presence of UL pilot contamination. The results highlighted the superior performance of the blind approach, especially when the channel hardening is low, being therefore useful to researchers working in the area of DL channel estimation. Future works include improving the blind estimation algorithm, for instance, finding a better switching point in Eq. (12), comparing the blind estimation benefit over the DL pilot-based method and evaluating these methods under other types of channel models, e.g., keyhole and Rician channels.

References

[1] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.
[2] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," in *Foundations and Trends in Signal Processing*. Hanover, MA, USA: Now Publ., 2021.
[3] 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, Nov. 2019.
[4] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, Jun. 2018.
[5] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, May 2017.
[6] P. Pasangi, M. Atashbar, and M. M. Feghhi, "Blind downlink channel estimation of multi-user multi-cell massive MIMO system in presence of the pilot contamination," *AEU-Int. J. Electron. Commun.*, vol. 117, Apr. 2020, Art. no. 153099.
[7] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," in *Foundations and Trends in Signal Processing*. Hanover, MA, USA: Now Publ., 2018.
[8] "Study on channel model for frequencies from 0.5 to 100 GHz, (release 16)," 3GPP, Sophia Antipolis, France, Rep. TR 38.901, 2019.
[9] "NR; user equipment (UE) radio transmission and reception; part 1: Range 1 standalone, (release 17)," 3GPP, Sophia Antipolis, France, Rep. TR 38.101-1, 2021.

Publication II

Souza, D. D., Freitas, M. M. M., Costa, D. B. da, Borges, G. S., Cavalcante, A. M., Valcarenghi, L., and Costa, J. C. W. A.
Effective Channel DL Pilot-Based Estimation in User-Centric Cell-Free Massive MIMO Networks

Reprinted with permission from
GLOBECOM 2022 - 2022 IEEE Global Communications Conference
pp. 705-710, 2022

Effective Channel DL Pilot-Based Estimation in User-Centric Cell-Free Massive MIMO Networks

Daynara D. Souza*, Marx M. M. Freitas*, Daniel B. da Costa†, Gilvan S. Borges*, André Mendes Cavalcante‡, Luca Valcarenghi⋆, and João C. Weyl Albuquerque Costa*
*Federal University of Pará, Belém, Brazil
†Technology Innovation Institute, Abu Dhabi, UAE
‡Ericsson Telecommunications S.A., Indaiatuba, Brazil
⋆Scuola Superiore Sant'Anna, Pisa, Italy
daynara@ufpa.br, marx@ufpa.br, daniel.costa@tii.ae, gilvan.borges@ifpa.edu.br,
andre.mendes.cavalcante@ericsson.com, luca.valcarenghi@santannapisa.it, jweyl@ufpa.br

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001150

***Abstract*—This paper investigates the performance of downlink (DL) pilot-based training to estimate the effective channel in user-centric cell-free massive multiple-input multiple-output (MIMO) networks. An algorithm for DL pilot assignment is proposed based on the level of interference between each user equipment (UE). It is proposed a refinement method for access point (AP) selection that controls the maximum AP cluster size of UEs. The strategy aims to control the maximum number of APs serving each UE to reduce the disparities among the AP cluster sizes. DL pilot-based training is compared with the blind, perfect and statistical channel state information (CSI) methods, assuming different precoding techniques, AP selection schemes, and the presence of pilot contamination. Our results demonstrate the following: (*i*) the proposed DL pilot assignment algorithm outperforms the baseline solutions; (*ii*) the proposed AP selection refinement method can improve the energy efficiency up to 86.6% without compromising the spectral efficiency; and (*iii*) DL pilot-based estimation reduces the normalized mean-square error significantly compared with blind and statistical CSI methods.**

***Index Terms*—AP selection refinement, cell-free massive MIMO networks, DL pilot-based estimation, DL pilot assignment, effective channel estimation, user-centric approach.**

I. Introduction

Cell-free (CF) massive multiple-input multiple-output (MIMO) networks consist of a large number of access points (APs) spread out in the coverage area, cooperating to serve the user equipments (UEs). Due to their distributed nature, these systems can provide increased macro-diversity and a more uniform spectral efficiency (SE) than the cell-based systems. The canonical version of CF considers that all APs serve all UEs. However, such a system demands enormous resource requirements (e.g., fronthaul signaling and processing) from the network, making it unscalable. [1], [2]. In this regard, the user-centric (UC) approach has emerged as an alternative to solve these drawbacks. By performing AP selection and limiting the number of UEs that each AP can serve, one can achieve scalability when the network resources (i.e., signal processing, signaling on fronthaul/backhaul, and total power) are independent of the number of UEs. However, strategies to control the size of AP clusters are still missing in the literature. The number of APs serving each UE can be very small or very large depending on the UE's position. In case of large AP clusters, the APs may become overloaded, impacting negatively the energy efficiency (EE) [3].

Another critical aspect of CF systems is that channel hardening may be less pronounced than in cellular systems due to the geographical distribution of the APs, with each one being equipped with a few antennas. The channel hardening phenomenon makes random channels behave almost deterministically when the number of antennas is large, enabling downlink (DL) data decoding based on statistical channel state information (CSI) of the effective channel [1], [2]. The low degree of hardening in CF systems may lead to the need for UEs to perform a more reliable channel estimation method. One of the main alternatives is the DL pilot-based estimation of the effective channel. This approach can be made more efficient by beamforming the DL pilots based on the uplink (UL) channel estimates at the APs, making the number of resource samples used on orthogonal pilot sequences a function of the number of UEs [4], [5]. However, a drawback of DL pilot-based estimation is that it has to use additional resource samples in training pilots, requiring small number of pilot sequences and, consequently, their reuse among the UEs.

This paper investigates the performance of the DL pilot-based estimation approach in UC CF massive MIMO networks, proposing novel algorithms for DL pilot assignment and cluster control for AP selection refinement. To the best of the authors' knowledge, such analysis has not been carried out in the literature yet. The analyzes are made considering different centralized and distributed precoding techniques. Numerical results are provided in terms of normalized mean square error (NMSE), SE, EE, and computational complexity (CC). The impact of different system configurations is investigated by varying some key parameters, such as coherence interval and pilot sequence length. Insightful discussions on the trade-off between estimation accuracy and overhead, and the performance of different estimation techniques are provided.

This work was supported by the Innovation Center, Ericsson Telecomunicações S.A, Brazil, by the National Council for Scientific and Technological Development (CNPq) and by the Coordination for the Improvement of Higher Education Personnel (CAPES).

The results show that the proposed DL pilot assignment can reduce the NMSE by 85%, while cluster control can improve the EE by 86.6%.

II. System Model

It is assumed a time-division duplex (TDD) CF massive MIMO system with M APs, equipped with N antennas each, and K single-antenna UEs, where $L = MN > K$. The TDD frame length is equal to the coherence interval. Hence, the channel is assumed to be static within a frame and but it varies independently for each frame. The channel between the AP m and UE k (and vice-versa), $\mathbf{h}_{mk} \in \mathbb{C}^{N\times 1}$, undergoes independent correlated Rayleigh fading, being defined as

$$\mathbf{h}_{mk} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \mathbf{R}_{mk}\right), \tag{1}$$

where $\mathbf{R}_{mk} \in \mathbb{C}^{N\times N}$ represents the covariance matrix modeling the large-scale fading behavior, considering spatial channel correlation, path loss, and shadowing.

The UL channels are estimated by correlating a received UL pilot signal with a corresponding known pilot sequence and performing minimum mean square error (MMSE) estimation. For each coherence interval of length τ_c (in symbols), all UEs simultaneously send UL pilot sequences of length τ_{up} samples. The pilot sequences are assumed to be pair-wisely orthogonal, and different UEs can be assigned to the same pilot sequence, i.e., they are reused when $K > \tau_{up}$.

In the DL, it is performed an AP selection scheme to determine the subset $\mathcal{M}_k \subset \{1, \ldots, M\}$ of APs serving UE k, which can also be represented by the diagonal matrix $\mathbf{D}_{mk} \in \mathbb{N}^{N\times N}$, i.e.,

$$\mathbf{D}_{mk} = \begin{cases} \mathbf{I}_N & m \in \mathcal{M}_k \\ \mathbf{0}_N & m \notin \mathcal{M}_k. \end{cases} \tag{2}$$

Then, each AP serves the UEs by implementing power control and precoding based on the UL estimates. Each AP only serves a limited number of UEs to address scalability aspects [1], [2]. Thus, it follows that $|\mathcal{D}_m| \leq \tau_p$, where $\mathcal{D}_m$ is a subset containing the UEs served by the AP m. Let $q_k(n)$ be the n-th symbol intended for UE k. It is assumed that $\mathbb{E}\left\{\mathbf{q}(n)\mathbf{q}(n)^H\right\} = \mathbf{I}_K$, where $\mathbf{q}(n) \triangleq [q_1(n), \cdots, q_K(n)]^T$. The data signal sent by AP m can be written as

$$\mathbf{x}_m(n) = \sum_{k=1}^{K} \mathbf{D}_{mk}\mathbf{w}_{mk}q_k(n), \tag{3}$$

where the term $\mathbf{w}_{mk} \in \mathbb{C}^{N\times 1}$ represents the precoding vector, such that $\mathbb{E}\left\{\left\|\mathbf{w}_{mk}\right\|^2\right\} = \bar{\rho}_{mk} = \rho_{mk}/\sigma_{dl}^2$, with ρ_{mk} being the transmit power that AP m assigns to the UE k and σ_{dl}^2 is the noise power. UE k receives a linear combination of the signals transmitted by the APs, i.e.,

$$\begin{aligned} y_{d,k}(n) &= \sum_{m=1}^{M} \mathbf{h}_{mk}\mathbf{x}_m(n) + n_{d,k}(n) \\ &= \underbrace{\alpha_{kk}q_k(n)}_{\text{desired signal}} + \underbrace{\sum_{k'\neq k}^{K} \alpha_{kk'}q_{k'}(n)}_{\text{inter-user interference}} + \underbrace{n_{d,k}(n)}_{\text{noise}}, \end{aligned} \tag{4}$$

where

$$\alpha_{kk'} = \sum_{m=1}^{M} \mathbf{h}_{mk}^{\mathrm{H}}\mathbf{D}_{mk'}\mathbf{w}_{mk'}, \quad k' = 1, \cdots, K. \tag{5}$$

The noise at the receiver follows a complex Gaussian distribution with zero mean and unit variance, i.e., $n_{d,k}(n) \sim \mathcal{CN}(0,1)$, α_{kk} is the effective channel for UE k, and $\alpha_{kk'}, k' \neq k$ is the effective interfering channel. In order to coherently detect the transmitted data symbol q_k, UE k should have sufficient knowledge of the effective channel α_{kk}, which can be acquired by estimation or statistical CSI [4]–[7].

III. AP Selection Refinement Method

This paper proposes a strategy to control the maximum AP cluster size of the UEs, by restricting the cardinality ($|\mathcal{M}_k|$) of the AP clusters to a limit called C_{max}, $\forall k \in \{1, \cdots, K\}$. The latter represents the maximum number of APs that each UE can connect. Let A_c denote the number of connections that the M APs (each serving at most τ_{up} UEs) can provide to the network. C_{max} can be calculated as $\mathrm{C}_{max} = \max(1, \alpha A_c/K)$, where $A_c = \tau_{up}M$, and $0 < \alpha \leq 1$ is a refinement parameter that modifies the stringency of C_{max}.

Therefore, when a UE is connected to an excessive number of APs (i.e., $|\mathcal{M}_k| \geq \mathrm{C}_{max}$) after performing AP selection, a central processing unit (CPU) is activated to drop the UE's connection with the APs having the weakest channel gains so that $|\mathcal{M}_k| = \mathrm{C}_{max}$. To this end, the CPU performs a sort operation in ascending order to identify the APs with the weakest channel gain to the UE k, with $\bar{\beta}_{m1} \leq \bar{\beta}_{m2} \leq \cdots \leq \bar{\beta}_{mk}$, where $\bar{\beta}_{mk}$ denotes the sorted version of $\beta_{mk} = \mathrm{tr}\{\mathbf{R}_{mk}\}/N$, $\forall m \in \mathcal{M}_k$. Then, it calculates $\mathrm{E}_k = |\mathcal{M}_k| - \mathrm{C}_{max}$, where E_k is the number of excessive APs in the AP cluster of the UE k. Let $\mathcal{E}_k$ denote the subset containing indexes of the first E_k APs presenting the lowest values in $\bar{\beta}_{mk}$. Thus, the CPU imposes that $\mathbf{D}_{mk} = \mathbf{0}_N$, $\forall m \in \mathcal{E}_k$. This policy aims to reduce the disparity among the cluster sizes of UEs by using the single refinement parameter α.

IV. DL Pilot-Based Channel Estimation and Pilot Assignment

When using DL pilot-based estimation scheme, the APs send DL pilots to the UEs by precoding them based on the UL channel estimates [4], [6]. The AP $m \in \mathcal{M}_{k'}$ precodes the DL pilot sequences $\boldsymbol{\psi}_{z_{k'}} \in \mathbb{C}^{\tau_{dp}\times 1}$, where τ_{dp} is the DL pilot length, such that the received DL pilot at the UE k can be written as

$$\mathbf{y}_{dp,k} = \sqrt{\tau_{dp}} \sum_{k'=1}^{K} \alpha_{kk'}\boldsymbol{\psi}_{z_{k'}} + \mathbf{n}_{dp,k}, \tag{6}$$

where $\mathbf{n}_{dp,k}$ is the noise vector whose elements are independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ random variables (RVs). It is assumed that the DL pilot sequences are pair-wisely orthonormal, i.e.,

$$\boldsymbol{\psi}_{z_1}^{\mathrm{H}}\boldsymbol{\psi}_{z_2} = \begin{cases} 1, \text{ if } z_1 = z_2 \\ 0, \text{ if } z_1 \neq z_2, \end{cases} \tag{7}$$

and are also reused among the UEs when $K > \tau_{dp}$. Accordingly, UE k correlates the received signal with a known pilot sequence $\boldsymbol{\psi}_{z_k}$ in order to estimate its effective channel α_{kk}, such that

$$\hat{y}_{dp,k} = \boldsymbol{\psi}_{z_k}^{\mathrm{H}} \mathbf{y}_{dp,k} = \sqrt{\tau_{dp}}\alpha_{kk} + \sqrt{\tau_{dp}} \sum_{k' \neq k}^{K} \alpha_{kk'} \boldsymbol{\psi}_{z_k}^{\mathrm{H}} \boldsymbol{\psi}_{z_{k'}} + n_{p,k}, \tag{8}$$

where $n_{p,k} = \boldsymbol{\psi}_{z_{k'}}^{\mathrm{H}} \mathbf{n}_{dp,k}$. Then, the UE k performs the linear MMSE estimation of its channel α_{kk}, which is given by

$$\hat{\alpha}_{kk} = \mathbb{E}\{\alpha_{kk}\} + \frac{\mathrm{Cov}\{\alpha_{kk}, \hat{y}_{dp,k}\}}{\mathrm{Cov}\{\hat{y}_{dp,k}, \hat{y}_{dp,k}\}} \left(\hat{y}_{dp,k} - \mathbb{E}\{\hat{y}_{dp,k}\}\right). \tag{9}$$

The second term in (8) contains the pilot contamination effect generated by the pilot-sharing UEs. Algorithm 1 presents a DL pilot assignment method that minimizes the pilot contamination interference, where the index of the pilot assigned to the UE k is denoted as $z_k \in \{1, \ldots, \tau_{dp}\}$. In this one, the first τ_{dp} UEs are assigned to orthogonal sequences. The remaining UEs are assigned to the pilot that causes the lowest pilot contamination, given by the average power of the DL pilot contamination term $\mathbb{E}\{|\alpha_{ki}|^2\}$. This method is similar to the UL pilot assignment proposed by [2], with the difference being the pilot contamination term.

Algorithm 1: DL pilot assignment that aims to minimize pilot contamination

Input: DL pilot length τ_{dp}.
for $k = 1, \ldots, \tau_{dp}$ **do**
  $z_k \leftarrow k$
end
for $k = \tau_{dp} + 1, \ldots, K$ **do**
  $\zeta \leftarrow \underset{z \in \{1,\ldots,\tau_{dp}\}}{\arg\min} \sum_{i=1, z_i = z}^{k-1} \mathbb{E}\{|\alpha_{ki}|^2\}$
  $z_k \leftarrow \zeta$
end
Output: Pilot assignment indexes $z_1, \ldots, z_K$

V. Normalized Mean Square Error, Spectral and Energy Efficiency

The performance of the DL pilot-based estimation method can be computed by the NMSE between the channel estimate $\hat{\alpha}_{kk}$ and the effective channel α_{kk}, i.e.,

$$\mathrm{NMSE}_k = \frac{\mathbb{E}\left\{|\alpha_{kk} - \hat{\alpha}_{kk}|^2\right\}}{\mathbb{E}\left\{|\alpha_{kk}|^2\right\}}. \tag{10}$$

The achievable DL SE for UE k based on the use-and-then-forget (UatF) lower bound can be computed as [4], [7]

$$\mathrm{SE}_k = \frac{\tau_d}{\tau_c} \log_2\left(1 + \mathrm{SINR}_k\right), \tag{11}$$

with

$$\mathrm{SINR}_k = \frac{\left|\mathbb{E}\left\{\frac{\alpha_{kk}}{\hat{\alpha}_{kk}}\right\}\right|^2}{\mathrm{Var}\left\{\frac{\alpha_{kk}}{\hat{\alpha}_{kk}}\right\} + \sum\limits_{k' \neq k}^{K} \mathbb{E}\left\{\left|\frac{\alpha_{kk'}}{\hat{\alpha}_{kk}}\right|^2\right\} + \mathbb{E}\left\{\frac{1}{|\hat{\alpha}_{kk}|^2}\right\}}, \tag{12}$$

where $\hat{\alpha}_{kk}$ is the estimate of the effective channel α_{kk} and SINR_k is the signal-to-interference-plus-noise ratio (SINR) of UE k. To account for channel estimation overhead, the SE is multiplied by a pre-log factor $\tau_d/\tau_c = (\tau_c - \tau_p)/\tau_c = 1 - \tau_p/\tau_c$, i.e., the fraction of samples used for DL data transmission, where τ_p is the total number of samples used for UL and DL pilot estimation.

The total EE in bit/Joule is defined as the ratio between the total throughput $R_t = B\sum_{k=1}^{K} \mathrm{SE}_k$ in bit/s, where B is the bandwidth in Hz, and the total power consumed by all APs in Watts, including the consumption of amplifiers, circuits and backhaul links connecting them to the CPU [3], i.e.,

$$\mathrm{EE}_t = \frac{R_t}{\sum\limits_{m=1}^{M} \left\{\frac{\sigma_{dl}^2}{\gamma_m} \mathbb{E}\left\{\|\mathbf{x}_m\|^2\right\} + N P_{\mathrm{tc},m} + P_{\mathrm{bh},m}\right\}}, \tag{13}$$

where $0 < \gamma_m \leq 1$ denotes the efficiency of the power amplifier, and $P_{\mathrm{tc},m}$ is the power required of each antenna of the AP m to run internal components, such as converters and filters. Additionally, $P_{bh,m}$ is the power that the backhaul link connecting the CPU and AP m consumes, given by $P_{\mathrm{bh},m} = P_{0,m} + P_{\mathrm{bt},m} B \sum_{k \in \mathcal{D}_m} \mathrm{SE}_k$, where $P_{0,m}$ is a fixed power consumption of each backhaul and $P_{\mathrm{bt},m}$ is the traffic-dependent power in Watt per bit/s.

VI. Numerical Results

In order to evaluate the network performance of the proposed solutions, Monte-Carlo simulations are run. The simulation scenario consists of $M = 100$ APs, each equipped with $N = 4$ antennas, and $K = 20$ UEs covering a 1 km^2 rectangular area. It is assumed that the UEs are uniformly distributed into the coverage area and that the APs are placed following a hard core point process (HCPP). The propagation model adopted for the simulations is the 3GPP Urban Micro (UMi) path-loss model in which the line-of-sight (LOS) condition uses the probability functions defined in 3GPP TR 38.901 [8]. The correlation matrices $\mathbf{R}_{mk}$ are computed using the UMi path-loss model and the local scattering spatial correlation model presented in [2]. The simulations parameters are presented in Table I. The parameters for EE are set as $\gamma_m = 0.4$, $P_{\mathrm{tc},m} = 0.2$ W, $P_{0,m} = 0.825$ W, and $P_{\mathrm{bt},m} = 0.25$ W/(Gbit/s) [3].

It is considered two types of network implementations: (i) distributed configuration, where each AP performs the channel estimation, precoding and power allocation locally, and (ii) centralized configuration, where these tasks are performed by the CPU [1]. For the distributed implementation, the power coefficients at AP m are set as $\rho_{mk} = \rho_d\sqrt{\beta_{mk}}/\sum_{k' \in \mathcal{D}_m} \sqrt{\beta_{mk'}}$, where ρ_d is the maximum DL transmit power. For the centralized one, it is used the scalable

fractional power control [2]. In order to compute the precoding vectors, it is employed the partial MMSE (P-MMSE) and partial regularized zero-forcing (P-RZF) for the centralized implementation. For the distributed, it is utilized the local partial MMSE (LP-MMSE) and maximum ratio (MR). These techniques were chosen due to their scalability features [1].

TABLE I: Parameters and models used in the simulations.

Parameter	Value
Effective environment height, h_E	1.0 m
Shadow fading standard deviation, σ_{SF}	4 dB
Antenna height AP, UE, h_{AP}, h_{UE}	11.65 m, 1.65 m
RX noise Figure, NF	8 dB
Coherence interval τ_c	200 samples
UL and DL pilot length	$\tau_{up} = \tau_{dp} = 10$ samples
Carrier Frequency f, Bandwidth	3.5 GHz, 100 MHz
AP's total DL power	23 dBm
UE's total UL power	20 dBm
Angular standard deviations (ASDs)	$\sigma_\varphi = \sigma_\theta = 15°$
Antenna spacing	1/2 wavelength distance

Fig. 1 shows the cumulative distribution function (CDF) of the NMSE of different DL pilot assignment methods. One can note that the proposed DL pilot assignment in Algorithm 1, which is a non-joint method that aims to minimize pilot contamination, performs better than the other ones, reducing the NMSE by 85% compared with the joint method in the 50$^{\text{th}}$ percentile. The joint UL and DL strategy assigns orthogonal DL pilots to those UEs that use the same UL pilots [4]. This implies that non-joint methods can improve NMSE, even though they do not consider the number of pilot sequences as a function of the number of UEs. Therefore, providing higher performance with the advantage of being scalable, differently from the joint strategies which require $\tau_{up}\,\tau_{dp} \geq K$.

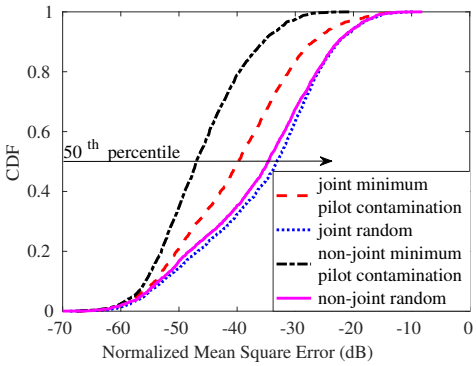


Fig. 1: CDF of the NMSE for different DL pilot assignment methods using MR precoding.

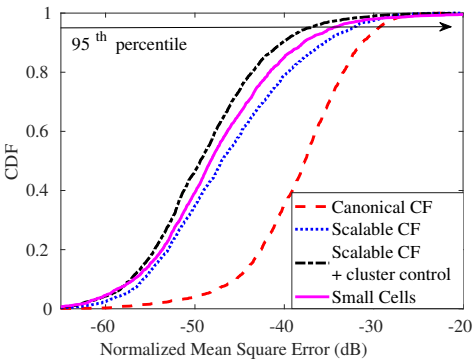
Fig. 2 shows the impact of AP selection in the DL pilot-based estimation. In Fig. 2a, one can note that AP selection techniques are crucial for reducing channel estimation errors since the NMSE is reduced in all AP clustering schemes compared to canonical CF. More specifically, the NMSE becomes even smaller in strategies that makes the UE connect to fewer APs, such as the proposed cluster control and small cells. The cluster control can reduce the NMSE up to 5 dB for the 95$^{\text{th}}$ percentile compared to the scalable CF strategy. Nonetheless, a solution like small cells (a single AP serving the UE, simulated by setting $C_{max} = 1$) can damage the SE as Fig. 2b depicts. One can note that even though it provides higher levels of EE in Fig. 2c, it does not provide a uniform coverage. This implies that the parameter α, employed in the cluster control strategy, has to be well fitted to avoid decreases in SE. For $\alpha = 25\%$, the cluster control can provide gains of up 19% to the 95$^{\text{th}}$ percentile of the SE for MR precoding and can improve the EE up to 86.6% for the P-MMSE. It happens because the cluster control provides the same SE while reducing the average number of APs serving each UE ($|\mathcal{M}_k|$) and the average number of UE served by an AP ($|\mathcal{D}_m|$), as Table II demonstrates, which also reduces the number of complex scalars exchanged with the fronthaul/backhaul and CC [1].

TABLE II: Average number of APs per UE ($|\mathcal{M}_k|$) and UEs per AP ($|\mathcal{D}_m|$). All standard deviations were around zero.

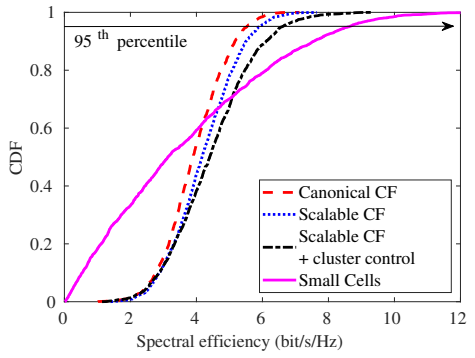
Method	mean $\lvert\mathcal{M}_k\rvert$	mean $\lvert\mathcal{D}_m\rvert$
Scalable CF	50	10
Scalable CF + cluster control	12	2.4

Fig. 3 evaluates the CDF of the NMSE and SE for MR and LP-MMSE precoding. The performance of DL pilot-based estimation is compared with blind estimation, statistical CSI, and perfect CSI. Blind estimation is a method that does not require time-frequency resource samples. It uses the average power of the received DL data signals to estimate the effective channel and is implemented following the steps presented in [9]. The conventional statistical CSI method uses average effective channel $\mathbb{E}\{\alpha_{kk}\}$ as the estimates, and an achievable DL SE can be computed using the *hardening bound* [1], [2]. The perfect CSI curves represent the SE when the UE has perfect knowledge of the effective channel, achieved in a genie-aided manner [2]. The estimation overhead is $\tau_p = \tau_{up} + \tau_{dp}$ for DL pilot-based estimation, whereas for the other estimation methods $\tau_p = \tau_{up}$, which is used in the pre-log factor of the SE. From Fig. 3, one can note that the DL pilot-based estimation decreases the NMSE compared with blind and conventional statistical CSI approaches for both precoding schemes. For instance, DL pilot-based estimation can reduce the NMSE by 95% and 71% compared with blind estimation for MR and LP-MMSE, respectively. From the SE results using MR precoding, it can be seen that blind and DL pilot-based estimation methods improve the system performance significantly compared with statistical CSI, up to 80.7% improvement for the 80% likely UEs. However, for LP-MMSE, DL pilot estimation degrades the SE due to the estimation overhead, whereas the improvement is slight for the blind approach. It is expected that the need to estimate the effective channel with LP-MMSE would be lower due to the higher degree of channel hardening.

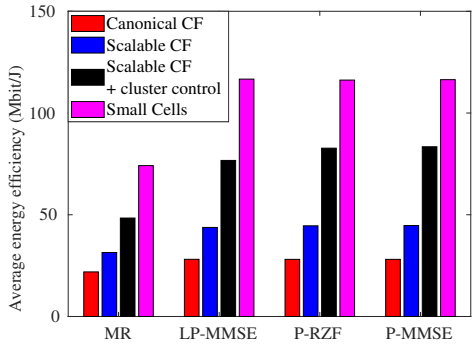
In Fig. 4, the coherence interval length, τ_c, is varied to evaluate the performance in terms of average SE for higher mobility and dispersion scenarios. One can note that blind



(a) NMSE of MR

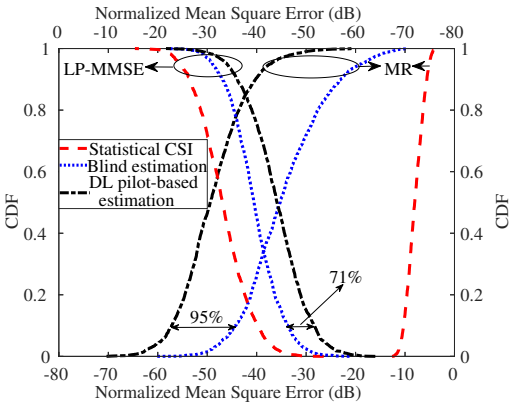


(b) SE of MR

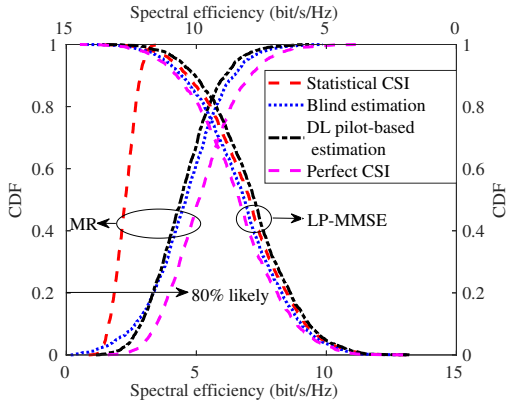


(c) EE

Fig. 2: CDF of the NMSE, SE and average total EE of DL pilot-based estimation for different AP selection schemes.



(a) NMSE



(b) SE

Fig. 3: CDF of the NMSE and SE for different estimation methods.

estimation is the best method for $\tau_c \leq 200$ samples, while DL pilot estimation has the best performance for $\tau_c \geq 300$ samples. The reason for that is the low pre-log factor value of the DL pilot estimation for the $\tau_c \leq 200$, although it has higher estimation accuracy. These results indicate that deciding which method is best for the UC CF scenario may vary depending on the system parameters.

To analyze the impact of the number of DL pilot sequences (τ_{dp}) in the performance of DL pilot-based method, Fig. 5 shows the average NMSE and SE versus τ_{dp}. As expected, the estimation accuracy improves as τ_{dp} increases since there is less pilot contamination interference. The SE for MR precoding can be maximized by setting $\tau_{dp} = 5$, which balances its estimation overhead and accuracy. On the other hand, the average SE decreases as the number of DL pilot sequences increases for the other precoding schemes, becoming best to set τ_{dp} at values as small as one, or performing another channel estimation method such as the blind one.

The CC of blind and DL pilot-based estimation methods can also be compared. Assuming that the statistical values are precomputed, known, and stay the same throughout communication, there is no need to evaluate its CC for each coherence interval. This also means that the CC for statistical CSI is zero. The CC of DL pilot-based estimation is a function of the pilot length, τ_{dp}, as each UE has to correlate its pilot sequence with the received pilot signal, an operation that requires $\tau_{dp} - 1$ additions and τ_{dp} multiplications. For blind estimation, each UE has to compute the average sample power of the received signal, its CC depends on the number of samples used for DL

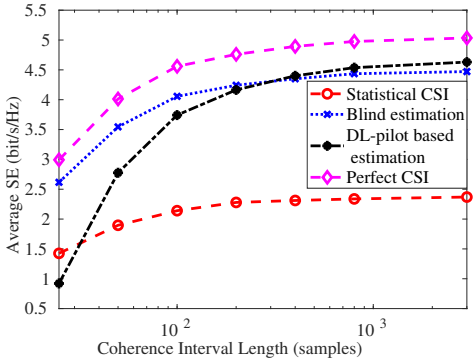
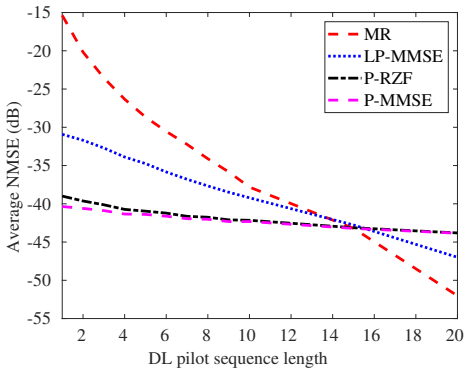
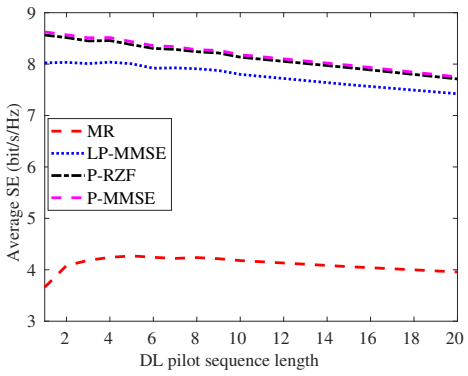


Fig. 4: Average SE versus coherence interval length.



(a) NMSE



(b) SE

Fig. 5: DL pilot-based estimation average NMSE and SE versus the DL pilot sequence length.

data transmission and requires $\tau_d - 1$ additions [9]. Table III summarizes these values, where it can be noted that their total CC is the same if $\tau_{dp} = \tau_d/2$. As typically $\tau_{dp} < \tau_d/2$, it is expected that the UEs using blind estimation will have higher CC than DL pilot-based estimation. Furthermore, there is also the CC of the proposed DL pilot assignment in Algorithm 1, which is $\mathcal{O}((K - \tau_{dp})\tau_{dp})$ for all UEs K, but the UEs do not perform this task.

TABLE III: CC for each UE by using different estimation methods, in every coherence interval.

Method	# additions	# multiplications	Total
DL pilot estimation	$\tau_{dp} - 1$	τ_{dp}	$2\tau_{dp} - 1$
Blind estimation	$\tau_d - 1$	0	$\tau_d - 1$

VII. Conclusions

This paper investigated the performance of DL pilot-based estimation in UC CF massive MIMO systems. It is proposed an algorithm for DL pilot assignment that aims to minimize pilot contamination. The paper also proposed a refinement method for AP selection that controls the maximum AP cluster size of UEs. The results demonstrate that the proposed DL pilot assignment algorithm can outperform the baseline solutions, reducing by 85% the NMSE, and has the advantage of being scalable. The proposed AP selection refinement method can improve the EE by up to 86.6% without compromising the SE. The results also demonstrated that for MR precoding, blind and DL pilot-based estimation methods can improve the system performance significantly compared to using only the statistical CSI. DL pilot-based estimation decreases NMSE and increases the SE of the 80% likely UEs by about 80.7% compared with statistical CSI. It is demonstrated that when the pilot overhead is small, DL pilot-based estimation is the best method since it has higher estimation accuracy. The analyzes indicate that it is possible to get the best performance with any one of the three estimation methods depending on the system parameters. This opens the way for future works to design self-regulated resource management strategies adapted for each scenario.

References

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

[2] Ö. Demir, E. Björnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO*, ser. Foundations and Trends in Signal Processing. Hanover, MA, USA: Now Publ., 2021.

[3] 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.

[4] 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.

[5] G. Interdonato, P. Frenger, and E. G. Larsson, "Utility-based downlink pilot assignment in cell-free massive MIMO," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, 2018, pp. 1–8.

[6] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, "How much do downlink pilots improve cell-free massive MIMO?" in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–7.

[7] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, 2017.

[8] "Study on channel model for frequencies from 0.5 to 100 GHz, (release 16)," 3GPP, Sophia Antipolis, France, Tech. Rep. TR 38.901, 2019.

[9] 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.

Publication III

Souza, D. D., Fernandes, A. L. P., Freitas, M. M. de, Costa, D. B. da, Cavalcante, A. M., Nardelli, P. H. J., and Costa, J. C. W. A.
Effective Channel Hybrid Estimation in User-Centric Distributed Massive MIMO Networks

Reprinted with permission from
2025 IEEE Wireless Communications and Networking Conference (WCNC)
pp. 1–6, 2025

Effective Channel Hybrid Estimation in User-Centric Distributed Massive MIMO Networks

Daynara D. Souza*†, André L. P. Fernandes*, Marx M. M. Freitas*, Daniel B. da Costa‡, André Mendes Cavalcante§, Pedro H. J. Nardelli†, and João C. W. A. Costa*
*Federal University of Pará, Belém, Brazil
†Lappeenranta-Lahti University of Technology, Lappeenranta, Finland
‡King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia
§Ericsson Telecommunications Ltda., Indaiatuba, Brazil
daynara.dias.souza@lut.fi, andrelpf@ufpa.br, marx@ufpa.br, danielbcosta@ieee.org, andre.mendes.cavalcante@ericsson.com, pedro.nardelli@lut.fi, jweyl@ufpa.br

2025 IEEE Wireless Communications and Networking Conference (WCNC) | 979-8-3503-6836-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/WCNC61545.2025.10978257

***Abstract*—User-centric (UC) distributed massive multiple-input multiple-output (D-mMIMO), commonly called cell-free mMIMO, is an essential technology for ensuring more uniform coverage and higher spectral and energy efficiencies in next-generation communication systems. This paper investigates an alternative effective channel estimation method to address the issue of the lower channel hardening degree experienced by UC D-mMIMO systems. Specifically, this paper proposes a hybrid approach for effective channel estimation, allowing users to estimate their channels either through downlink pilot-based training, by using a blind algorithm, or by relying on statistical channel state information (CSI). The hybrid estimation method is compared with the case where each method is applied individually, as well as with the ideal case of perfect CSI. The analysis is conducted using various system parameter settings, such as the number of antennas and users, and it accounts for the presence of pilot contamination. Simulation results reveal that the proposed hybrid channel estimation algorithm is able to provide the best spectral efficiency performance in any network parameter configuration, while reducing the estimation normalized mean-square error compared to conventional statistical CSI.**

***Index Terms*—Blind estimation, cell-free massive MIMO, channel hardening, downlink pilot-based estimation, distributed massive MIMO, effective channel estimation, user-centric approach.**

I. Introduction

Distributed massive multiple-input multiple-output (D-mMIMO) networks, also referred to as cell-free massive MIMO, are envisioned as one of the promising technologies for next-generation wireless networks. Unlike traditional systems, in which user equipments (UEs) connect to a specific cell tower, it allows UEs to communicate with multiple antennas spread across different locations [1]. From the UEs' perspective, their performance is consistent regardless of their location in the coverage area, effectively addressing the severe cell-edge interference of traditional systems, thereby enabling high data rates and simultaneous service for multiple UEs [2]. As the D-mMIMO scenario spreads many transmission and reception points (TRPs) over the coverage area to serve the UEs, it is essential to define the subset of TRPs serving each UE properly. For this purpose, it is employed a user-centric (UC) approach to achieve a scalable and efficient system, such that the complexity and resource requirements for each TRP must remain manageable even as the number of UEs tends towards infinite [3].

Despite the benefits, a limiting aspect of UC D-mMIMO systems is that the channel hardening may be less pronounced than in cell-based systems due to the TRPs being scattered across different locations as well as by the fact that the UEs connect only to a subset of TRPs equipped with few antennas [4]–[6]. Essentially, the channel hardening property makes random channels behave almost deterministically when the number of antennas is large, allowing for downlink (DL) data decoding based on statistical channel state information (CSI) of the effective channel [3]. Nonetheless, the low degree of hardening in UC D-mMIMO systems may lead to the need for UEs to perform a more reliable method for effective channel estimation [7]–[9].

In [9], blind and DL pilot-based estimation methods for UC D-mMIMO are investigated. The results revealed that both approaches can improve the system performance significantly compared to using only the statistical CSI. However, these channel estimation methods have their advantages and drawbacks. For instance, using DL pilot estimation requires time-frequency resources that increase the transmission overhead, while blind estimation generally has lower accuracy. As the spectral efficiency (SE) depends on estimation accuracy and overhead, UEs may favor one channel estimation method over the other depending on the network configuration.

Therefore, this paper proposes a hybrid channel estimation method to dynamically choose the estimation method according to the available resources and the UE channel condition. In this approach, each UE estimates its effective channel using the conventional method called statistical CSI, blind estimation, or DL pilot-based training. To the best of the authors' knowledge, this is the first paper that proposes a hybrid estimation method to balance overhead, SE, and estimation normalized mean square error (NMSE). Moreover, this paper presents insightful discussions on which estimation method is more appropriate for a given situation in UC D-mMIMO networks under correlated Rician fading channels. As far as the authors are aware, this is also the first work that caries out performance analysis of effective channel estimation methods under Rician channels. The numerical results demonstrate the effectiveness of the hybrid method, showing that it can strike a balance between estimation accuracy and overhead to provide the best SE performance.

Notation: Boldface lowercase and uppercase letters denote

979-8-3503-6836-9/25/$31.00 ©2025 IEEE

column vectors and matrices, respectively, the superscript $(\cdot)^{\mathrm{H}}$ denotes the conjugate-transpose operation, the $N \times N$ identity matrix is $\mathbf{I}_N$, and the cardinality of the set $\mathcal{A}$ is represented by $|\mathcal{A}|$. The trace, euclidean norm and expectation operator are denoted as $\mathrm{tr}(\,.\,)$, $\|\,.\,\|$ and $\mathbb{E}\{\,.\,\}$, respectively, and the notation $\mathcal{N}_{\mathbb{C}}\left(\mu, \sigma^2\right)$ stands for a complex Gaussian random variable with mean μ and variance σ^2.

II. System Model

We consider a D-mMIMO network composed of L TRPs and K single-antenna UEs, where $L > K$. Each TRP is equipped with N antennas, and the total number of antennas considering all TRPs is $M = NL$. The TRPs connect to the central processing units (CPUs) through fronthaul links, while the CPUs are linked to each other through backhaul ones. The system operates on time-division duplex (TDD) mode and it is assumed reciprocity for the uplink (UL) and DL channels.

The channel vector $\mathbf{h}_{kl} \in \mathbb{C}^{N\times 1}$ between the TRP l and UE k undergoes an independent correlated Rician fading, being defined as [5], [6], [10]

$$\mathbf{h}_{kl} = \underbrace{\sqrt{\frac{\kappa_{kl}\beta_{kl}}{1+\kappa_{kl}}}\mathbf{h}_{kl}^{\mathrm{LOS}}}_{\overline{\mathbf{h}}_{kl}} e^{j\psi_{kl}} + \underbrace{\sqrt{\frac{\beta_{kl}}{1+\kappa_{kl}}}\mathbf{h}_{kl}^{\mathrm{NLOS}}}_{\widetilde{\mathbf{h}}_{kl}}, \quad (1)$$

where $\overline{\mathbf{h}}_{kl} \in \mathbb{C}^{N\times 1}$ means the deterministic line-of-sight (LOS) component and $\widetilde{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \widetilde{\mathbf{R}}_{kl}) \in \mathbb{C}^{N\times 1}$ is the small-scale fading from non-line-of-sight (NLOS) component with covariance matrix $\widetilde{\mathbf{R}}_{kl} = \mathbb{E}\{\widetilde{\mathbf{h}}_{kl}\widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} \in \mathbb{C}^{N\times N}$. Due to the UEs mobility, random phase shifts $\psi_{kl} \sim \mathcal{U}[0, 2\pi)$ occur in the LOS component, which is assumed to be unknown. The power ratio between the LOS and NLOS components is denoted by the Rician factor κ_{kl}, and β_{kl} represents the large-scale fading, encompassing path loss and shadowing.

Each coherence block comprises τ_c samples, where τ_p samples are dedicated for UL pilot signals and τ_d for DL data. In the UL training phase, the UEs send pilot sequences of τ_p-length to the TRPs. Then, the UL channels are estimated using phase-unaware linear minimum mean square error (LMMSE) estimation. The pilot signals are orthogonal to each other, and a pilot t_k can be reused by some UEs if $K > \tau_p$. Let $\mathcal{P}_k \subset \{1, \ldots, K\}$ denote the subset of the UEs assigned to the pilot t_k, including the UE k. The received pilot signal at TRP l can be expressed as [3]

$$\mathbf{y}_{t_k l}^{\mathrm{pilot}} = \sum\nolimits_{i\in\mathcal{P}_k} \sqrt{\tau_p \eta_i}\, \mathbf{h}_{il} + \mathbf{n}_{t_k l}, \quad (2)$$

where $\mathbf{n}_{t_k l} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_N, \sigma_{ul}^2 \mathbf{I}_N\right)$ denotes the noise and η_i is the power that the UE i transmits in the UL direction. The LMMSE channel estimate is given by $\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p \eta_k}\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_k l}^{-1}\mathbf{y}_{t_k l}^{\mathrm{pilot}}$, where $\mathbf{R}_{kl} = \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\mathrm{H}}\} = (\overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\mathrm{H}} + \widetilde{\mathbf{R}}_{kl})$ and $\boldsymbol{\Psi}_{t_k l} = \mathbb{E}\{(\mathbf{y}_{t_k l}^{\mathrm{pilot}})(\mathbf{y}_{t_k l}^{\mathrm{pilot}})^{\mathrm{H}}\} = \sum_{i\in\mathcal{P}_k} \eta_i \tau_p (\overline{\mathbf{h}}_{il}\overline{\mathbf{h}}_{il}^{\mathrm{H}} + \widetilde{\mathbf{R}}_{il}) + \sigma_{\mathrm{ul}}^2 \mathbf{I}_N$.

In UC systems, each UE is associated with a subset of TRPs called TRP cluster, represented by $\mathcal{M}_k \subset \{1, \ldots, L\}$. The connections between the UE k and TRPs are denoted by a diagonal matrix $\mathbf{D}_{kl} \in \mathbb{N}^{N\times N}$, being defined as

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & \text{if } l \in \mathcal{M}_k \\ \mathbf{0}_N & \text{if } l \notin \mathcal{M}_k. \end{cases} \quad (3)$$

The subset of UEs served by an TRP is denoted by $\mathcal{D}_l$. We consider $|\mathcal{D}_l| \leq \tau_p$ and $|\mathcal{M}_k| \leq C_{\max}$ as restrictions to ensure system scalability and reduce processing complexity [3], [9], [11]. Let $s_k \in \mathbb{C}$ denote the symbol intended for the UE k and $\mathbf{x}_l = \sum_{k=1}^{K} \mathbf{D}_{kl}\mathbf{w}_{kl}s_k$ represent the data signal sent by the TRP l, where $\mathbf{w}_{kl}$ denotes the precoding vector. The DL received signal at the UE k can be expressed as

$$y_k^{dl} = \sum_{l=1}^{L} \underbrace{\mathbf{h}_{kl}^{\mathrm{H}}\mathbf{D}_{kl}\mathbf{w}_{kl}s_k}_{\text{Desired signal}} + \sum_{i\neq k}^{K}\sum_{l=1}^{L} \underbrace{\mathbf{h}_{kl}^{\mathrm{H}}\mathbf{D}_{il}\mathbf{w}_{il}s_i}_{\text{Interfering signals}} + \underbrace{n_k}_{\text{Noise}}, \quad (4)$$

where $n_k \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma_{\mathrm{dl}}^2\right)$ is the receiver noise. The terms s_k and $\mathbf{w}_{kl}$ satisfy $\mathbb{E}\left\{\|s_k\|^2\right\} = 1$ and $\mathbb{E}\left\{\|\mathbf{w}_{kl}\|^2\right\} = \rho_{kl}$, with ρ_{kl} being the power allocated from TRP l to the UE k.

III. Downlink Effective Channel Estimation

The *effective channel* is a scaling factor containing the channel distortion on the data symbol received by a given UE. The effective channel can be essentially defined as the summation of the inner product between the channel vector $\mathbf{h}_{kl}$ and the precoding vector $\mathbf{D}_{il}\mathbf{w}_{il}$ from all its serving TRPs. Then, the DL effective channel gains are given by

$$\alpha_{ki} = \sum\nolimits_{l=1}^{L} \mathbf{h}_{kl}^{\mathrm{H}}\mathbf{D}_{il}\mathbf{w}_{il}, \quad i = 1, \cdots, K. \quad (5)$$

where α_{kk} is called the effective channel for UE k and α_{ki} is the effective interfering channel, with $i \neq k$. In order to coherently detect the transmitted data symbol s_k, UE k should have sufficient knowledge of the effective channel α_{kk}, which can be acquired by estimation or statistical CSI [7], [12].

The effective channel value can remain nearly constant even if the individual elements of the L channel vectors $\{\mathbf{h}_{k1}, \cdots, \mathbf{h}_{kL}\}$ are changing. Specifically, when the random realizations of the effective channel are close to the mean value (i.e., small variance), then approximately the same effective channel appears in every coherence block, being almost like communicating over a deterministic channel [3]. However, the different geographical locations of the TRPs may lead to different large-scale fading coefficients for each TRP, which is the main reason why the channel hardening degree may be lower in D-mMIMO systems than in cellular ones.

A. Statistical CSI (sCSI)

In the statistical CSI approach, UEs do not estimate their effective channel; they rely on the hardening property and decode their received signal based on the average value of the effective channel. This approach is most employed in the massive MIMO literature [3], [13]. In this case, the effective channel estimate can be set as

$$\hat{\alpha}_{kk}^{\mathrm{sCSI}} = \mathbb{E}\{\alpha_{kk}\}. \quad (6)$$

B. DL pilot estimation (DLPE)

One alternative approach is to use DL pilot-based estimation scheme, in which the TRPs send DL pilots to the UEs by precoding them based on the UL channel estimates [7], [9]. In this method, the TRP $l \in \mathcal{M}_k$ precodes the DL pilot sequences $\boldsymbol{\psi}_{z_k} \in \mathbb{C}^{\tau_{dp}\times 1}$ for UE k, where τ_{dp} is the DL pilot length. It is assumed that the DL pilot sequences are reused among the UEs when $K > \tau_{dp}$ and that they are pair-wisely orthonormal, i.e., $\boldsymbol{\psi}_{z_1}^{\mathrm{H}}\boldsymbol{\psi}_{z_2}$ equals one if $z_1 = z_2$ and zero otherwise. Accordingly, UE k correlates the received signal with a known pilot sequence $\boldsymbol{\psi}_{z_k}$ in order to estimate its effective channel α_{kk}, such that

$$\hat{y}_k^{dp} = \sqrt{\tau_{dp}}\alpha_{kk} + \sqrt{\tau_{dp}}\sum\nolimits_{i\neq k}^{K}\alpha_{ki}\boldsymbol{\psi}_{z_k}^{\mathrm{H}}\boldsymbol{\psi}_{z_i} + n_{p,k}, \quad (7)$$

where $n_{p,k} \sim \mathcal{CN}(0,\sigma_{dl}^2)$. The second term in (7) contains the pilot contamination effect generated by the pilot-sharing UEs. Then, UE k performs the LMMSE estimation of its channel α_{kk}, which is given by

$$\hat{\alpha}_{kk}^{\mathrm{DLPE}} = \mathbb{E}\{\alpha_{kk}\} + \frac{\mathrm{Cov}\{\alpha_{kk},\hat{y}_k^{dp}\}}{\mathrm{Cov}\{\hat{y}_k^{dp},\hat{y}_k^{dp}\}}\left(\hat{y}_k^{dp} - \mathbb{E}\{\hat{y}_k^{dp}\}\right). \quad (8)$$

C. Blind estimation (BE)

It is also possible to estimate the effective channels gains without using any DL pilots, a method called blind estimation [8], [12]. The first step of the blind estimation algorithm is to compute the sample average power of the received signal at UE k per coherence interval, i.e.,

$$\xi_k = \frac{\sum_{n=1}^{\tau_d}\left|y_k^{dl}(n)\right|^2}{\tau_d}, \quad (9)$$

where $y_k^{dl}(n)$ denotes the n-th sample received at UE k, shown in (4), and $\tau_d = \tau_c - \tau_{up}$ is the number of symbols per coherence interval used for DL data transmission. Then, the effective channel blind estimator is formulated as [8]

$$\hat{\alpha}_{kk}^{\mathrm{BE}} = \begin{cases}\sqrt{\xi_k - Z_k}, & \text{if } \xi_k > Z_k,\\ \mathbb{E}\{\alpha_{kk}\}, & \text{otherwise,}\end{cases} \quad (10)$$

where

$$Z_k = \sum\nolimits_{i=1,i\neq k}^{K}\mathbb{E}\left\{\left|\alpha_{ki}\right|^2\right\} + \sigma_{dl}^2. \quad (11)$$

We assume that the UE knows $\mathbb{E}\{\alpha_{kk}\}$ and Z_k since they are functions only of the large-scale coefficients, which are constant over many coherence intervals [3], [8], [12].

D. NMSE and Achievable SE

The NMSE between the channel estimate $\hat{\alpha}_{kk}$ and the effective channel α_{kk} can be computed as

$$\mathrm{NMSE}_k = \frac{\mathbb{E}\left\{\left|\alpha_{kk} - \hat{\alpha}_{kk}\right|^2\right\}}{\mathbb{E}\left\{\left|\alpha_{kk}\right|^2\right\}}. \quad (12)$$

The achievable DL SE when UE k is performing some method to estimate the effective channel can be computed based on the use-and-then-forget (UatF) lower bound, which is written as [7], [12]

$$\mathrm{SE}_k^{\mathrm{eCSI}} = \frac{\tau_d}{\tau_c}\log_2\left(1+\mathrm{SINR}_k^{\mathrm{eCSI}}\right), \quad (13)$$

with

$$\mathrm{SINR}_k^{\mathrm{eCSI}} = \frac{\left|\mathbb{E}\left\{\frac{\alpha_{kk}}{\hat{\alpha}_{kk}}\right\}\right|^2}{\mathrm{Var}\left\{\frac{\alpha_{kk}}{\hat{\alpha}_{kk}}\right\} + \sum\limits_{i\neq k}^{K}\mathbb{E}\left\{\left|\frac{\alpha_{ki}}{\hat{\alpha}_{kk}}\right|^2\right\} + \sigma_{dl}^2\mathbb{E}\left\{\frac{1}{|\hat{\alpha}_{kk}|^2}\right\}}, \quad (14)$$

where $\hat{\alpha}_{kk}$ is the effective channel estimate. Other approaches could be used to compute the achievable SE, e.g., the methods in [7, Sec. III-A] and [12, Sec. V], but the first one is not a rigorous lower bound while the second one requires complicated numerical computations, unlike the UatF method.

For comparison purposes, it can also be derived a SE expression for the theoretical case where the UEs know their effective channels perfectly, obtained in some genie-aided manner. The SE for the perfect CSI is given by [3], [7]

$$\mathrm{SE}_k^{\mathrm{pCSI}} = \frac{\tau_d}{\tau_c}\mathbb{E}\left\{\log_2\left(1+\mathrm{SINR}_k^{\mathrm{pCSI}}\right)\right\}, \quad (15)$$

where

$$\mathrm{SINR}_k^{\mathrm{pCSI}} = \frac{\left|\alpha_{kk}\right|^2}{\sum_{\substack{i=1\\ i\neq k}}^{K}\left|\alpha_{ki}\right|^2 + \sigma_{\mathrm{dl}}^2}. \quad (16)$$

IV. Effective Channel Hybrid Estimation in Distributed Massive MIMO Networks

This section presents the proposed method for performing hybrid channel estimation, in which the UEs can estimate their channel by using either DL pilot estimation, blind estimation, or statistical CSI. Fig. 1 illustrates the proposed hybrid DL channel estimation strategy. The DL channel estimation decision starts after the UL channel estimation phase, so the CSI of UE k is known. The CPU computes a decision metric to define which estimation method has the best performance for UE k based on the CSI. The decision metric is a function of the pre-log factor to account for estimation overhead, SE, and NMSE for each method, such as

$$\mathrm{DM}_k = p_f\left[w\frac{\mathrm{SE}_k}{\mathrm{SE}_k^{\mathrm{pCSI}}} + (1-w)(1-\mathrm{NMSE}_k)\right], \quad (17)$$

with

$$p_f = \frac{\tau_d}{\tau_c} = \begin{cases}(\tau_c - \tau_{up})/\tau_c & \text{for sCSI and BE}\\ (\tau_c - \tau_{up} - \tau_{dp})/\tau_c & \text{for DLPE}\end{cases}, \quad (18)$$

where $\mathrm{SE}_k^{\mathrm{pCSI}}$ is the achievable SE for perfect channel estimation, which is used to normalize the SE term, and $w \in [0,1]$ is a weight between the SE and NMSE terms. The NMSE and SE values are computed using (12) and (13), respectively. Moreover, the estimation overhead is taken into account by setting the pre-log factor as in (18). That is, the estimation overhead for statistical CSI and blind estimation is $\tau_p = \tau_{up}$, while it is computed as $\tau_p = \tau_{up} + \tau_{dp}$ for DL pilot estimation.

To compute the decision metric for statistical CSI, named $\mathrm{DM}_k^{\mathrm{sCSI}}$, the effective channel estimate is set as in (6). The decision metric for DL pilot-based estimation is computed by employing (8), and is denoted as $\mathrm{DM}_k^{\mathrm{DLPE}}$. The decision

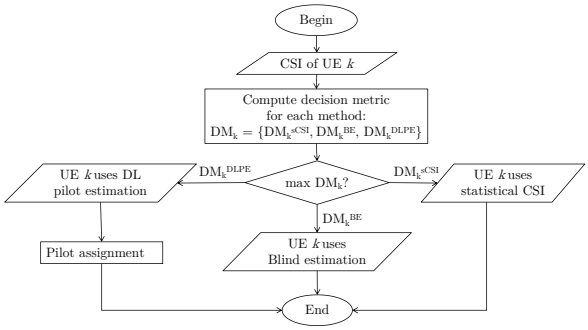


Fig. 1. Proposed hybrid DL channel estimation strategy. In this scheme, the CPU uses the UL channel estimates to compute the decision metric. Afterwards, it informs the estimation method each UE uses for DL data decoding, which is the one that maximizes the decision metric.

metric for blind estimation method is denoted as $\mathrm{DM}_k^{\mathrm{BE}}$ and is computed using (10). Then, the method with the highest value of (17) is chosen for each UE.

Furthermore, the UEs estimating the channel through DL pilot-based are assigned to a DL pilot of length $\bar{\tau}_{dp} = \min\{\tau_{dp}, |\mathcal{A}|\}$, where $\tau_{dp} < \tau_c$ is the maximum DL pilot length and $\mathcal{A} \subset \{1, \ldots, K\}$ is the subset of UEs using the DL pilot-based estimation. To avoid data-to-pilot contamination, $\bar{\tau}_{dp}$ samples cannot be used for data by the UEs estimating their channels by blind estimation or statistical CSI. Therefore, after the proposed hybrid strategy computation finishes, the pre-log factor at the SE for all UEs is

$$p_f^{\text{Hybrid}} = \frac{\tau_c - \tau_{up} - \bar{\tau}_{dp}}{\tau_c}. \tag{19}$$

V. Numerical Results

This section presents illustrative numerical examples to show the efficiency of the proposed hybrid estimation method in UC D-mMIMO networks. The main simulation parameters are $L = 100$ TRPs, with each one being equipped with $N = 2$ antennas, and $K = 20$ UEs. It is considered $\tau_c = 200$, $\tau_{up} = 10$, $\tau_d = 190$, and $C_{\text{max}} = 10$. The maximum transmission powers of the UEs and TRPs are $100\,\mathrm{mW}$ and $200\,\mathrm{mW}$, respectively. It is assumed that the UEs are uniformly distributed into the square coverage area of $1\ \mathrm{km} \times 1\ \mathrm{km}$ and the TRPs are placed following a hard core point process (HCPP)[1]. The results are obtained through 500 Monte-Carlo simulations to account for different TRP/UE locations, each with 100 different channel realizations. These simulation parameters are in accordance with [3], [4], [7]. In the results, the decision metric for the hybrid estimation strategy in (17) is computed by setting $w = 0.5$. Moreover, some of the system's parameters are varied throughout the results to validate the analyzes, e.g., the coherence interval length, the number of antennas per TRP, and the number of UEs. Furthermore, we have considered both centralized and distributed network implementation using maximum ratio (MR), local partial minimum mean square error (MMSE) (LP-MMSE), and partial MMSE (P-MMSE) schemes for precoding the DL data.

[1]The TRPs are initially placed randomly based on a homogeneous Poisson point process with mean rate $1/d_{\text{min}}$, where $d_{\text{min}} = \sqrt{A/L}$ is the minimum distance between TRPs and A is the coverage area. Then, TRPs that do not meet the spacing requirement are randomly replaced.

TABLE I
Parameters assumed for the path loss and local scattering spatial correlation model.

Parameter	Value
Shadow fading standard deviation, σ_{SF}	4 dB
TRP/UE antenna height, $h_{\text{TRP}}, h_{\text{UE}}$	11.65 m, 1.65 m
RX noise figure (NF)	8 dB
Carrier frequency, bandwidth	3.5GHz, 100MHz
Angular standard deviations (ASDs)	$\sigma_\varphi = \sigma_\theta = 15^\circ$
Antenna spacing	1/2 wavelength distance

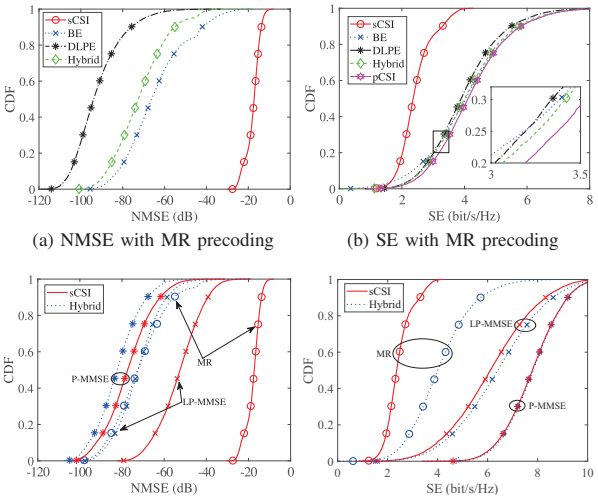
It is used a joint UL pilot assignment and TRP selection scheme presented in [3, Sec. 4.4]. In this method, the first τ_{up} UEs are assigned mutually orthogonal pilots, and the remaining UEs are assigned to the pilot that causes the lowest pilot contamination. After that, each TRP chooses to serve τ_{up} UEs with the largest channel gain in each pilot. The DL pilot assignment is performed similarly, but minimizing the term $\mathbb{E}\{|\alpha_{ki}|^2\}$, which is the average power of DL pilot contamination [9]. For power allocation, the power coefficients are directly proportional to the large-scale fading gains in distributed implementation, that is, $\rho_{kl} = \rho_d \sqrt{\beta_{kl}} / \sum_{k' \in \mathcal{D}_l} \sqrt{\beta_{k'l}}$, where ρ_d is the maximum DL transmit power per TRP and $\beta_{kl} = \mathrm{tr}(\mathbf{R}_{kl})/N$. For the centralized one, it is used the scalable fractional power control with the following parameters: $v' = -0.5$ and $\kappa' = 0.5$ [3]. The shadowing terms of an TRP to different UEs are correlated and the covariance matrices follow the local scattering spatial correlation model, both computed as in [3]. The propagation model adopted is the 3GPP Urban Micro path loss model with LOS and NLOS conditions defined in 3GPP TR 38.901 [14]. Table I specifies the parameters used as entries for the path loss and covariance models [3], [15].

Fig. 2 presents the cumulative distribution function (CDF) of NMSE and SE of each effective channel estimation method. It can be seen in Fig. 2a that the proposed hybrid method has a higher NMSE than the DL pilot estimation. This happens because the decision metric depends on a weighing between NMSE and SE also the pre-log factor. Then, the overhead of DL pilot estimation may be too high that choosing an estimation method with higher NMSE and lower overhead is the best approach. From the CDFs of SE in Fig. 2b, it can be observed that blind estimation and DL pilot estimation significantly outperform statistical CSI. This finding indicates that relying on channel hardening in UC D-mMIMO systems can be detrimental even when considering the Rician fading channel model. This results is noteworthy since Rician fading incorporates a LOS component, which typically enhances channel hardening compared to Rayleigh fading [5]. The plausible explanation is that the LOS component is less prevalent in larger coverage areas due to the increased distance between UEs and TRPs. As a result, channel hardening is less pronounced in outdoor scenarios considered in this paper. Moreover, it can be noticed in Fig. 2b that the CDFs of blind

estimation and DL pilot estimation cross each other, but the hybrid method outperforms all estimation methods when used individually. These results in Figs. 2a and 2b indicate that the decision metric with weighting coefficient $w = 0.5$ is able to balance NMSE and SE.

In Figs. 2c and 2d, NMSE and SE performances of the hybrid strategy are also shown for LP-MMSE and P-MMSE precoding. Noticeable, unlike with MR, the performance gains are less pronounced. This is expected since these precoding choices enhance channel hardening [6], [8]. However, these results indicate that the hybrid strategy can also improve estimation accuracy and SE of other precoding schemes, when compared with statistical CSI.

It is also interesting to evaluate the frequency, in percentage, that each estimation method is chosen for the UEs after computing the hybrid strategy, which is computed as the ratio of UEs assigned to each estimation method by the total number of UEs, across all Monte-Carlo simulations. This result can be seen in Table II for different precoding schemes. For MR, most UEs (86.2%) perform blind estimation, while DL pilot-based estimation is performed by 13.78% of the cases, and statistical CSI is rarely selected as the best approach for the UEs. For LP-MMSE and P-MMSE, the preference for blind estimation is even higher, and both statistical CSI and DL pilot estimation are rarely or not selected as the best approach for any UE. This preference for blind estimation is due to the fact that the slight NMSE and SE improvements do not compensate for the high estimation overhead when using DL pilot estimation.



(a) NMSE with MR precoding (b) SE with MR precoding

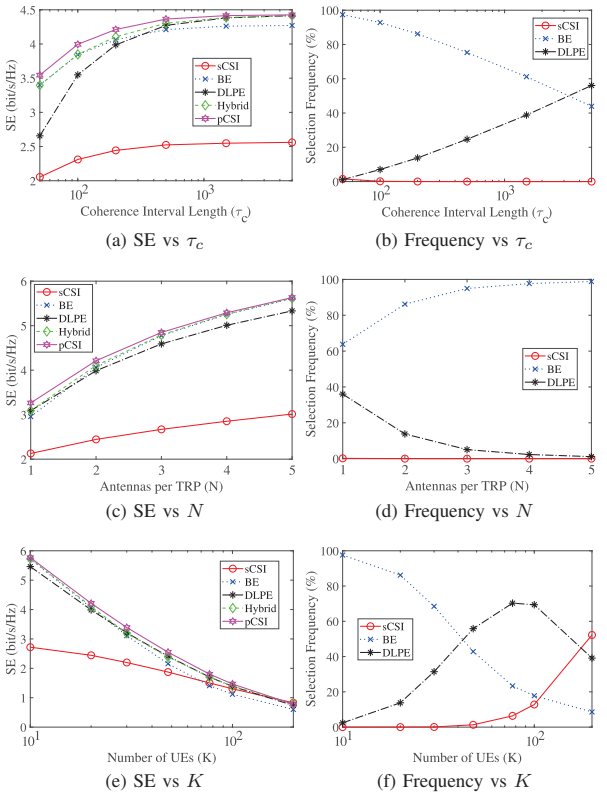
(c) NMSE with different precoding (d) SE with different precoding

Fig. 2. CDF of the NMSE and SE for different estimation methods.

Fig. 3 portraits the average SE when the coherence interval length, the number of antennas per TRP, and the number of UEs are varied. In these results, only one parameter is varied per time, while the remaining parameters are kept the same as in previous analyses. To analyze how the hybrid strategy dynamically adapts for the UEs' needs for any circumstance,

TABLE II
FREQUENCY (%) THAT ESTIMATION METHOD IS SELECTED IN THE HYBRID STRATEGY FOR DIFFERENT PRECODING SCHEMES.

Precoding	Estimation Method		
	sCSI	BE	DLPE
MR	0.02	86.2	13.78
LP-MMSE	0.05	99.95	0
P-MMSE	0	100	0



(a) SE vs τ_c (b) Frequency vs τ_c

(c) SE vs N (d) Frequency vs N

(e) SE vs K (f) Frequency vs K

Fig. 3. Average SE and frequency that estimation method is selected in the hybrid strategy, both as a function of the system parameters with MR precoding.

Fig. 3 also presents the frequency in percentage that each estimation method is chosen for the UEs after computing the hybrid strategy by varying the system's parameters.

In Fig. 3a, the coherence interval length is varied from $\tau_c = 50$ to $\tau_c = 5000$ to illustrate the average SE performance for different scenarios, where having more samples indicates lower mobility and dispersion scenarios. One can note in Fig. 3a that the blind estimation is the best method for $\tau_c \leq 200$ samples, while DL pilot estimation has the best performance for $\tau_c \geq 400$ samples. The reason for that is the low pre-log factor value (higher overhead) of the DL pilot estimation for the $\tau_c \leq 200$, although it has higher accuracy. For the estimation selection frequency in Fig. 3b, it can be noticed that for lower values of τ_c, blind estimation is performed by most UEs. As τ_c increases, the percentage of UEs using

blind estimation decreases and DL pilot-based increases, being almost 50%-50% in $\tau_c = 3000$. This happens because the overhead of DL pilot-based decreases as τ_c increases, allowing more UEs to select DL pilot-based estimation.

Fig. 3c exhibits the average SE versus the number of antennas per TRP, varying from $N = 1$ to $N = 5$. Fig. 3c shows that increasing N improves the SE in all estimation schemes. It can be noted that the blind estimation method has the most significant increase in SE as N varies. For lower values of N, DL pilot-based estimation has better performance than blind estimation, while the inverse happens for higher N. Once again, the hybrid strategy can achieve the best performance for any value of N. Furthermore, as opposed to the variation of τ_c, the percentage of UEs using blind estimation increases with N while the DL pilot-based percentage decreases. This happens because the channel hardening enhances when N increases, consequently improving the estimation accuracy of blind estimation, thus more UEs select this method. From these results, it is clear that the hybrid strategy also self-regulates the choice of estimation method each UE should perform as the number of antennas per TRP N varies.

In Fig. 3e, the average SE is plotted as a function of the number of UEs, where K varies from 10 to 200. It can be noticed that the best benchmark estimation method changes depending on the number of UEs. However, as in Fig. 3c, the results indicate that the hybrid strategy can provide the best SE in the UC D-mMIMO with any configuration setup. Finally, the estimation method selection frequency when varying the number of UEs is shown in Fig. 3f, where it can be noted that there is a different behavior from the previous parameters. The estimation method performed by most UEs changes three times within the considered values. Another difference is that for a large number of UEs K, the statistical CSI method begins to be used by the UEs, even being more frequent than the blind estimation approach. This shows that although statistical CSI is infrequent in most situations, it should still be considered by the proposed hybrid strategy as one of the possible estimation methods for the UEs. In a nutshell, these results demonstrate that the hybrid method provides an objective manner to reduce the pre-log factor of the DL pilot-based estimation by allowing other UEs to use methods such as blind or statistical CSI, according to their needs.

VI. Conclusions

This paper presented the results of the proposed hybrid method for effective channel estimation in UC D-mMIMO systems. The proposed strategy is an approach in which the UEs can estimate their channel using either DL pilot estimation, blind estimation, or statistical CSI. The hybrid approach is motivated by the fact that each estimation method can be the best for different situations, e.g., UE channel condition, total transmit power, and available time-frequency resources. Then, the hybrid strategy is performed to provide the best performance in any situation. The hybrid approach uses a decision metric to determine the most suitable method for each UE. Therefore, it is a self-regulated resource management strategy that can adapt to the available network resources, the system's parameters, and the channel conditions. The results were analyzed in terms of NMSE and achievable DL SE for different precoding schemes under correlated Rician fading channels. Moreover, the coherence interval length, the number of antennas per TRP, and the number of UEs were varied to assess the proposed strategy's performance under different network conditions. The results demonstrated that the hybrid strategy can strike a balance between estimation accuracy and overhead to provide the best SE results. Future works include the expansion of this analysis for scenarios considering the mobility of UEs, evaluating both the impact of channel aging on the SE performance and also how to perform energy-efficient resource allocation and pilot assignment in such scenarios.

Acknowledgment

This paper was partly funded by EU MSCA project CO-ALESCE (n.101130739), EU CEF project CAREWINGS-5G (n.101180459), and by Research Council of Finland via X-SDEN (n.349965) and ECO-NEWS (n.358928). This work was also partly supported by CNPq, CAPES, and the Innovation Center, Ericsson Telecomunicações Ltda., Brazil.

References

[1] 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.

[2] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[3] Ö. Demir, E. Björnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO*. Foundations and Trends in Signal Processing Series, Now Publishers, 2021.

[4] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, 2018.

[5] A. A. Polegre, F. Riera-Palou, G. Femenias, and A. G. Armada, "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, 2020.

[6] T. T. R. Ueoka, D. D. Souza, M. M. d. Freitas, A. M. Cavalcante, and J. W. Costa, "Performance of cell-free massive MIMO networks under Rayleigh and Rician fading," in *SBrT 2022*, 2022.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] M. M. M. Freitas, D. D. Souza, A. L. P. Fernandes, 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. 1–7.

[12] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, 2017.

[13] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*. Foundations and Trends in Signal Processing Series, Now Publishers, 2018.

[14] 3GPP, *Study on channel model for frequencies from 0.5 to 100 GHz*, 2019, 3GPP TR 38.901 (Release 16).

[15] 3GPP, *NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone*, 2021, 3GPP TR 38.101-1 (Release 17).

Publication IV

Souza, D. D., Freitas, M. M. M., Fernandes, A. L. P., Nardelli, P. H. J., Costa, D. B. da, Cavalcante, A. M., and Costa, J. C. W. A.

Trajectory Optimization in User-Centric Distributed Massive MIMO Systems Enabled by UAV Swarms

Reprinted with permission from
IEEE Transactions on Vehicular Technology
Vol. 74, no. 6, pp. 9252 – 9268, June 2025

Trajectory Optimization in User-Centric Distributed Massive MIMO Systems Enabled by UAV Swarms

Daynara D. Souza, *Graduate Student Member, IEEE*, Marx M. M. Freitas, André L. P. Fernandes, Pedro H. J. Nardelli, *Senior Member, IEEE*, Daniel Benevides da Costa, *Senior Member, IEEE*, André Mendes Cavalcante, *Member, IEEE*, and João C. Weyl Albuquerque Costa, *Senior Member, IEEE*

***Abstract*—User-centric (UC) distributed massive multiple-input multiple-output (D-mMIMO), commonly called cell-free mMIMO, is an important technology to ensure a more uniform coverage and higher spectral and energy efficiencies in next-generation communication systems. This paper investigates the performance of UC D-mMIMO systems enabled by a swarm of uncrewed aerial vehicles (UAVs). Specifically, it presents a comprehensive study on the deployment and trajectory optimization of UAVs as aerial transmission and reception points (TRPs) of D-mMIMO systems, considering systems composed solely of aerial TRPs and those implemented combining aerial and terrestrial TRPs. Moreover, user equipment (UE) mobility is modeled using a discrete-time Markov chain, and a novel approach to heuristically optimize the positions of aerial TRPs is proposed, considering the continuous movement of UEs in the coverage area. The proposed approach optimizes the three-dimensional location of each UAV under a time discretization framework, with the positioning of the UAVs being adjusted periodically, allowing for iterative trajectory optimization to improve the spectral efficiency (SE) performance of the UEs. Simulation results reveal that the proposed UAV trajectory optimization allows for significant SE improvement, especially for a low UE density scenario. Specifically, comparing the proposed method with a fixed position setup, an up to 47.84% increase in average SE is achieved.**

***Index Terms*—Aerial TRPs, cell-free massive MIMO networks, computational complexity, deployment and trajectory optimization, user-centric approach, autonomous aerial vehicles.**

I. Introduction

USER-CENTRIC (UC) distributed massive multiple-input multiple-output (D-mMIMO) networks, also called cell-free (CF)-mMIMO networks, are envisioned as promising technologies for next-generation wireless networks. In D-mMIMO systems, many transmission and reception points (TRPs) are spread out in the coverage area, each equipped with one or more antennas [1], [2]. The various users' equipments (UEs) share the same time-frequency resources, and each UE is served by a specific subset of TRPs defined by their needs and requirements, giving rise to the UC concept. These networks stand out for their high spectral efficiency (SE), reduced susceptibility to blocking and shadowing, and uniform performance among UEs [3], [4].

Along with these advantages, new challenges also arise. Problems related to power control, signal precoding, resource allocation, decoding strategies, and interference management are critical issues. Furthermore, to achieve a scalable and efficient D-mMIMO system, the complexity and resource requirements for each TRP must remain finite when the number of UEs tends to infinity [5], [6], [7].

Nevertheless, deploying terrestrial infrastructure might not be financially or physically viable in practical situations, particularly in complex terrains and private or remote areas. Moreover, in some occasions, terrestrial wireless networks may be unavailable due to natural disasters, power outages, maintenance, and other factors. In those situations, aerial communication systems using uncrewed aerial vehicles (UAVs) present a promising new paradigm for the rapid and flexible deployment of communication infrastructure, owing to their high maneuverability [2], [8]. UAVs can be regarded as a reusable aircraft designed to operate without onboard crew, which can be controlled remotely or programmed to fly autonomously [9].

For each UAV, there are physical constraints on weight, size, and energy consumption that limit the amount of onboard

Received 3 December 2024; accepted 1 February 2025. Date of publication 5 February 2025; date of current version 20 June 2025. This work was supported in part by the Innovation Center, Ericsson Telecomunicações Ltda., Brazil, in part by the National Council for Scientific and Technological Development, in part by the Coordination for the Improvement of Higher Education Personnel, in part by the Research Council of Finland through (a) X-SDEN Project n.349965, (b) EnergyNet Fellowship n.321265/n.328869, and (c) ECO-NEWS Project n.358928, and in part by Jane and Aatos Erkko Foundation through STREAM Project, in part by EU MSCA "COALESCE" Project n.101130739, and in part by European Commission CEF CAREWINGS-5G Project n.101180459. The review of this article was coordinated by Dr. Cunhua Pan. *(Corresponding author: Daynara D. Souza.)*

Daynara D. Souza is with the School of Energy Systems, Lappeenranta-Lahti University of Technology, 53850 Lappeenranta, Finland, and also with the Applied Electromagnetism Laboratory, Federal University of Pará, Belém 66075-110, Brazil (e-mail: daynara.dias.souza@lut.fi).

Marx M. M. Freitas is with the Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, 03043 Cassino, Italy, and also with the Applied Electromagnetism Laboratory, Federal University of Pará, Belém 66075-110, Brazil (e-mail: marx@ufpa.br).

André L. P. Fernandes is with the Applied Electromagnetism Laboratory, Federal University of Pará, Belém 66075-110, Brazil (e-mail: andrelpf@ufpa.br).

Pedro H. J. Nardelli is with the School of Energy Systems, Lappeenranta-Lahti University of Technology, 53850 Lappeenranta, Finland (e-mail: Pedro.Nardelli@lut.fi).

Daniel Benevides da Costa is with the Interdisciplinary Research Center for Communication Systems and Sensing, Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia (e-mail: danielbcosta@ieee.org).

André Mendes Cavalcante is with the Ericsson Research, Ericsson Telecomunicações Ltda., Indaiatuba 13337-300, Brazil (e-mail: andre.mendes.cavalcante@ericsson.com).

João C. Weyl Albuquerque Costa is with the Applied Electromagnetism Laboratory, Federal University of Pará, Belém 66075-110, Brazil, and also with the Instituto Sustentabilidade da Amazônia com Ciência e Inovação, Belém 66620-270, Brazil (e-mail: jweyl@ufpa.br).

Digital Object Identifier 10.1109/TVT.2025.3539120

equipment to be carried, including the number of TRP antennas. Nonetheless, the gains of employing a large number of antennas can still be achieved by the cooperation of UAVs as the TRPs of a UC D-mMIMO system, even with a small number of antennas per TRP. In this case, the UAVs operate like a cooperative swarm that maintains and adjusts their formation to engage in collaborative missions by sharing their goals and individual energy and data resources. The UAV swarms can be particularly useful in scenarios where the dynamic nature of the wireless network is taken into account, such as when the UEs positions are constantly changing due to their mobility. The maneuverability of UAVs allows for their positions to be adjusted according to the current state of the system, and their trajectories can change along with the movement of UEs.

In this context, implementing a swarm of UAVs working collaboratively as TRPs for UC D-mMIMO networks has to be appropriately investigated, and several operational challenges must be addressed, such as data transmission in areas where a traditional network is not available or unable to support the demands. Additionally, the horizontal positions, altitudes, and trajectories of UAVs have to be optimized to adapt to the continuous movement of UEs. Therefore, deeper investigations into this topic are required because state-of-the-art solutions rely on the UEs position knowledge and an assumption that they remain fixed throughout the optimization process [10], [11], [12], [13], [14], [15], [16], [17], [18], [19].

A. Literature Review

1) UC D-mMIMO: Jointly processing signals from different TRPs is a solution proposed to work around the wireless networks with increasing interference. In the literature, this type of operation is referred to by different names, including network multiple-input multiple-output (MIMO), multi-cell MIMO cooperative networks, coordinated multipoint with joint transmission (CoMP-JT), and distributed MIMO.

Fundamentally, the co-processing in these systems can be implemented in two fashions: network- or user-centric. In the network-centric approach, the TRPs are divided into disjoint clusters, and each cluster transmits jointly to the UEs residing in their coverage area. However, this approach can limit performance gains at cluster borders because UEs may not be served by their optimal TRP set [3], [5].

On the other hand, the UC approach ensures that UEs receive transmissions from an optimized subset of TRPs that maximize performance. This subset considers TRP connection limits, channel conditions, and fronthaul capacity for signal sharing. This strategy virtually eliminates cell boundaries, offering superior interference management compared with the network-centric method [3], [5]. In [20], the maximum number of TRPs serving the UEs was limited and adjusted to avoid significant losses in SE and reduce computational complexity.

Although the architecture of UC D-mMIMO is fundamentally distributed, the processing implementation can be either centralized or distributed. In the distributed implementation, the TRPs perform tasks such as channel estimation, power control, TRP selection, and precoding computation. This approach aligns with the distributed nature of UC D-mMIMO and offers high computational resource efficiency. Conversely, these tasks are performed at edge central processing units (CPUs) in centralized processing, which enables more advanced processing techniques, potentially achieving higher performance at the cost of increased computational complexity [6]. However, centralized processing can also be scalable and have lower fronthaul signaling than the distributed case [6], [21].

Recent developments in UC D-mMIMO research have delved into its practical aspects. This includes considerations on transitioning from traditional cellular massive MIMO infrastructures and the integration into practical specifications and guidelines, such as the open radio access network (O-RAN) architecture [22], and also a comprehensive techno-economic study of this system under distributed and centralized signal processing implementations [23].

Therefore, the combining effort of several works dedicated to co-processing techniques has culminated in the development of the UC D-mMIMO system, proving its suitability as one of the leading solutions for next-generation wireless networks [3], [5], [6], [7], [20], [21], [22], [23]. Combining the benefits of the UC D-mMIMO system with aerial TRPs is an excellent opportunity to boost the performance gain of worst UEs while requiring low data transmission power in each UAV. However, a thorough study of how the cooperation of UAV swarms can achieve these performance gains is required. Moreover, there is a need for heuristic strategies because only these types of solutions can fulfill the scalability requirements [5].

2) UC D-mMIMO Enabled by UAVs: Recently, the integration of UAVs into wireless communication systems has been proposed to enhance network coverage, capacity, and energy efficiency. The case of D-mMIMO systems enabled by a collaborative UAV swarm has not yet been fully investigated in detail. In the works that leverage UAVs and D-mMIMO networks, UAVs primarily serve as either aerial TRPs (base stations) [10], [11], [12], [13], [14], [15], [16], [17], [18], [19] or aerial mobile UEs [24], [25]. In what follows, we provide insights into the latest works that leverage UAVs as TRPs of D-mMIMO communication networks, pointing out their strengths and weaknesses.

In [10], a framework was introduced for radio-frequency energy harvesting that combines D-mMIMO, UAVs, and reconfigurable intelligent surfaces. Such an innovative approach aims to provide a seamless energy source to remote Internet of Things (IoT) devices, reducing the required infrastructure of a high-density IoT network. However, it is noteworthy that the framework primarily emphasizes energy harvesting, with SE as a secondary objective. This approach makes sense because IoT devices generally require a low bandwidth. Moreover, the work lacks optimization for positions and trajectories of UAVs.

In [11], the authors proposed the utilization of high-amplitude platforms (HAPs) as a CPU to aggregate signals from UAVs through terahertz wireless fronthaul links. The study provided a promising deployment scheme to increase the data rates of UEs. However, it lacked optimization of the positions and heights of UAVs, focusing on the power optimization aspects of the system.

In contrast, [12] did not focus on HAPs, but it optimized the altitude and horizontal coordinates of UAVs to maximize the downlink (DL) rates of the UEs. Despite this, the latter study overlooked factors such as UEs mobility, trajectory optimization, and the possible variation of the height of each UAV. In addition, it assumed the Rayleigh fading channel model, with the effect of line-of-sight (LoS) and non-line-of-sight (NLoS) propagation used only to calculate the average channel gain. Lastly, it only analyzed the performance of conjugate beamforming, also called maximum ratio (MR).

In [13], an investigation of uplink (UL) communication in D-mMIMO networks utilizing UAVs as flying base stations was performed. The study primarily focused on an algorithmic approach based on asymptotic signal-to-interference-plus-noise ratio (SINR) expressions for a high number of UEs to perform a deployment optimization that maximizes the UL rates of the UEs. In [14], the impact of the imperfect wireless fronthaul link between the UAVs and the CPU was taken into account to evaluate the SE, allowing a more robust deployment optimization. A closer examination of possible UAV deployment strategies for D-mMIMO was presented in [15]. Despite this, [13], [14], [15] neglected the UEs mobility and trajectory optimization of the UAVs.

Furthermore, [16] proposed a hybrid system alternating between UAV swarms and satellites, focusing on radio resource allocation. However, the assumption of an arbitrarily given UAV trajectory limits the applicability of the proposed solution. This was addressed in [17], which proposed a trajectory optimization for the same scenario. However, both works relied on scenarios that leverage hybrid UAV and satellite systems to provide service for remote IoT devices. Therefore, they did not tackle the issue of trajectory optimization considering the UEs mobility.

In [18], the deployment of UAVs in areas without terrestrial infrastructure was explored, extending wireless resources to numerous service providers. However, the scenario in the work was a conventional cellular network instead of a UC D-mMIMO system. Moreover, the term "cell-free" that has been traditionally used for UC D-mMIMO was actually used to refer to an area without cell-based internet access. Similarly, the term "cell-free" is used without connection to UC D-mMIMO systems in [19], where each UEs was served by one UAV and had exclusive time-frequency resources.

Finally, in [24], [25], it was assumed that UAVs are the mobile UEs of the UC D-mMIMO system. In [24], power control and security aspects were investigated, while [25] evaluated whether the system can support UAV communications in the presence of interfering ground UEs. Although their proposals provide an interesting line of research, they follow a different investigation scenario than our work.

In summary, while the integration of UAVs into D-mMIMO communication systems holds considerable promise for enhancing wireless networks, current research often neglects crucial aspects such as UEs mobility, trajectory optimization, and heterogeneous deployment scenarios. Addressing these limitations is essential for realizing the full potential of UAV-enabled D-mMIMO communication systems in future wireless networks. It is noteworthy that most works optimize the UAV position assuming that the positions of the UEs are fixed and known at the aerial TRPs and/or CPU. However, trajectory optimization methods that can be applied for scenarios where UEs positions are unknown are still missing.

B. Contributions

By bridging some gaps in the literature, the primary objective of this paper is to formulate strategies for the deployment and trajectory optimization of UAVs in the context of UC D-mMIMO networks, where a swarm of UAVs cooperatively serve the UEs as aerial TRPs. This paper analyzes not only UC D-mMIMO systems consisting exclusively of aerial TRPs carried out by UAVs, but also a setup with mixed of terrestrial and aerial TRPs. The UEs mobility across the coverage area is modeled, and the optimization of the UAV trajectories is accomplished by dynamically adjusting the UAV positions while the UEs positions change with time. Additionally, the proposed algorithm uses the DL SE as a metric that links the UAV positions with the UEs performances. In summary, the algorithm first selects the UAVs that most need to update their positions. Then, each UAV initially moves in a random direction, and the SE is reevaluated. A new movement direction is computed from a truncated normal distribution, which gives a higher probability of either continuing in the same direction or turning in the opposite direction. This decision depends on whether the previous action led to an improvement in SE or not. This paper also leverages a closed-form expression for MR precoding to compute SE, thus requiring only slow-time varying statistical information and simplifying the optimization process. To the best of the authors' knowledge, this is the first paper proposing an approach to heuristically optimize the positions and trajectories of aerial TRPs in UC D-mMIMO networks, taking into account the constant movement of UEs in the coverage area. Moreover, the proposed method is designed to be scalable, and it does not require knowledge of the UEs positions at any moment by the TRPs or the CPU. Overall, the main contributions of this paper can be summarized as:

- UC D-mMIMO systems enabled by a swarm of UAVs are analyzed. Specifically, this paper models the DL SE assuming Rician fading, which takes into account the effect of LoS and NLoS propagation. In addition, the modeling of computational complexity (CC), fronthaul quantization, power consumption, and energy efficiency (EE) in such systems is presented, assuming the relevant parameters related to UAV communication systems.
- A new method for optimizing the trajectory of UAVs is proposed. This method optimizes the horizontal and vertical locations of each UAV. Moreover, the paper offers a model for UEs mobility using a discrete-time Markov chain. The proposed trajectory optimization takes into account the constantly changing positions of UEs. By deploying a time discretization framework, the positioning of UAVs can be adjusted at each time frame, allowing for iterative optimization of the UAV trajectories.

- In addition to the scenarios consisting exclusive of aerial TRPs carried by UAVs, this paper evaluates hybrid scenarios integrating both aerial and terrestrial TRPs to address challenges associated with the demands of high UEs density traffic and data rate.
- Four precoding strategies are analyzed in terms of the SE: MR, local partial MMSE (LP-MMSE), partial regularized zero-forcing (P-RZF), and partial MMSE (P-MMSE). The first two are implemented in a distributed fashion, and the two latter in a centralized fashion. The results show that the proposed algorithm is able to adjust the UAV positions with precoding schemes other than MR, even though it is based on simpler MR closed-form expressions for SE.
- The impact of the number of UEs and aerial TRPs on the gains achievable with the proposed method is analyzed for different signal processing alternatives. Simulation results reveal that the proposed solution allows for an SE improvement of 29.76% for MR, 47.84% for LP-MMSE, 38.14% for P-RZF, and 35.24% for P-MMSE, compared with the case where each UAV chooses a random initial position and stays fixed while only the UEs move.

C. Paper Outline and Notations

The remainder of this paper is organized as follows. Section II presents the system model, including the channel modeling and estimation procedure, the DL SE, and the CC to perform signal processing. Section III presents the modeling of power consumption and energy efficiency. Section IV introduces the proposed approach to optimize the trajectories of the UAVs. Section V plots illustrative numerical results and provides insightful discussions to reveal the effectiveness of the proposed approach. Finally, Section VI concludes the paper.

Notation: Boldface lowercase and uppercase letters denote column vectors and matrices, respectively, the superscript $(\,.\,)^{\mathrm{H}}$ denotes the conjugate-transpose operation, the $N \times N$ identity matrix is $\mathbf{I}_N$, and the cardinality of the set $\mathcal{A}$ is represented by $|\mathcal{A}|$. The trace, euclidean norm, and expectation operators are denoted as $\mathrm{tr}(\,.\,)$, $\|\,.\,\|$, and $\mathbb{E}\{\,.\,\}$, respectively. The floor, ceiling, and modulus operators are denoted as $\lfloor\,.\,\rfloor$, $\lceil\,.\,\rceil$, and $\%$, respectively. The notation $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ stands for a complex Gaussian random variable with mean μ and variance σ^2. $\mathcal{U}[a, b]$ denotes a uniform random variable in the interval $[a, b]$, and $\mathcal{F}(x, \sigma^2, a) \sim \mathcal{TN}(x, \sigma^2, x-a, x+a)$ stands for a truncated normal distribution function in the interval $[x-a, x+a]$ with the mean value centered at x degrees and variance σ^2. Moreover, $P(X = x|Y = y)$ denotes the conditional probability of $X = x$ given $Y = y$, where X and Y are random variables.

II. System Model

We consider a UC D-mMIMO network composed of L TRPs and K single-antenna UEs. Each TRP is equipped with N antennas, and the total number of antennas considering all TRPs is $M = NL$, where $M > K$. The aerial TRPs are deployed at UAVs, and they connect to the CPU through dedicated wireless fronthaul links, as illustrated in Fig. 1 a. The fronthaul links undergo a limited transmission capacity using analog-to-digital converters (ADCs) to quantize signals before being sent to the fronthaul. In the hybrid D-mMIMO system, terrestrial and aerial TRP are deployed, as depicted in Fig. 1 b. By denoting L_t and L_a as the number of terrestrial and aerial TRPs, the total number of TRPs is defined as $L = L_t + L_a$. Additionally, each UEs is served by a subset of TRPs. The system operates in the time-division duplex (TDD) mode, and reciprocity is assumed for the UL and DL channels. The channel $\mathbf{h}_{kl} \in \mathbb{C}^{N\times 1}$ between TRP l and UEs k undergoes an independent correlated Rician fading, defined as [25], [26], [27], [28]

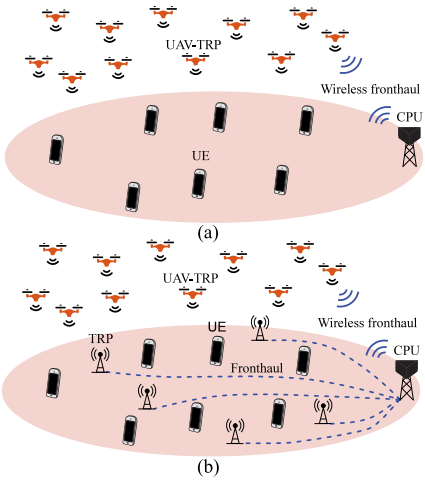


Fig. 1. UC D-mMIMO system enabled by a swarm of UAVs as aerial TRPs. (a) Only aerial TRPs. (b) Terrestrial and aerial TRPs.

$$\mathbf{h}_{kl} = \overline{\mathbf{h}}_{kl} e^{j\psi_{kl}} + \widetilde{\mathbf{h}}_{kl}, \tag{1}$$

$$\overline{\mathbf{h}}_{kl} = \sqrt{\frac{\kappa_{kl}\beta_{kl}}{1+\kappa_{kl}}} \left[1, \ldots, e^{-\mathrm{j}(N-1)\pi \sin(\varphi_{kl})\cos(\theta_{kl})}\right]^{\mathrm{T}}, \tag{2}$$

$$\widetilde{\mathbf{h}}_{kl} = \sqrt{\frac{\beta_{kl}}{1+\kappa_{kl}}} \mathbf{h}_{kl}^{\mathrm{NLoS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \widetilde{\mathbf{R}}_{kl}), \tag{3}$$

where $\overline{\mathbf{h}}_{kl} \in \mathbb{C}^{N\times 1}$ and $\widetilde{\mathbf{h}}_{kl} \in \mathbb{C}^{N\times 1}$ represent the weighted channel components associated with the LoS and NLoS propagation effects, respectively, and $\psi_{kl} \sim \mathcal{U}[0, 2\pi)$ indicates random phase shifts in the LoS component due to the UEs mobility. The channel $\overline{\mathbf{h}}_{kl}$ is computed assuming TRPs equipped with a half-wavelength-spacing uniform linear array (ULA), where φ_{kl} denotes the azimuth angle and θ_{kl} the elevation angle of the LoS component. Moreover, $\mathbf{h}_{kl}^{\mathrm{NLoS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{kl}^{\mathrm{NLoS}}) \in \mathbb{C}^{N\times 1}$, where the covariance matrix $\mathbf{R}_{kl}^{\mathrm{NLoS}}$ can be computed following the local scattering model for spatial covariance presented in [29, Sec. 2.6]. Thus, $\widetilde{\mathbf{R}}_{kl} = \mathbb{E}\{\widetilde{\mathbf{h}}_{kl}\widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} = \mathbf{R}_{kl}^{\mathrm{NLoS}}\beta_{kl}/(\kappa_{kl}+1) \in \mathbb{C}^{N\times N}$, and $\mathbf{R}_{kl} = \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\mathrm{H}}\} = (\overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\mathrm{H}} + \widetilde{\mathbf{R}}_{kl})$. Then, the large-scale fading gain can be calculated as $\beta_{kl} = \mathrm{tr}(\mathbf{R}_{kl})/N$, encompassing both path loss and shadowing. Furthermore, κ_{kl} denotes the Rician factor that represents the power ratio between the LoS and NLoS components. For LoS propagation links, the Rician factor can be modeled as $\kappa_{kl} = \min\{10^{1.3-0.003d_{lk}}, p_{\mathrm{LoS}}/(1-p_{\mathrm{LoS}})\}$ [25], [28], where d_{lk} is

the distance between TRP l and UEs k, and p_{LoS} is the LoS propagation probability. For aerial TRPs, p_{LoS} is a function of the elevation angle θ_{kl}, given by [30]

$$p_{\text{LoS}} = \frac{1}{1 + a \exp\left(-b\left[\frac{180\theta_{kl}}{\pi} - a\right]\right)}, \tag{4}$$

where a and b are determined according to the environment. The large-scale fading β_{kl} is modeled as [30]

$$\beta_{kl}[\text{dB}] = -20\log\left(\frac{4\pi f_c d_{kl}}{c}\right) - \begin{cases} \eta^{\text{LoS}} \text{ for LoS} \\ \eta^{\text{NLoS}} \text{ for NLoS} \end{cases}, \tag{5}$$

where the first term denotes the free-space path loss, which is a function of the distance d_{kl} between UEs k and TRP l, the carrier frequency f_c, and the speed of light c. The variables η^{LoS} and η^{NLoS} stand for the additional losses that differ according to the environment, and the presence or absence of the LoS component.

A. Uplink Training and Channel Estimation

Each coherence block comprises τ_c samples, where τ_p samples are dedicated to UL pilot signals and τ_d for DL data. In the UL training phase, the UEs send pilot sequences of τ_p-length to the TRPs. Then, the UL channels are estimated using phase-unaware linear minimum mean square error (LMMSE) estimation. The pilot signals are orthogonal to each other, and a pilot t_k can be reused by some UEs if $K > \tau_p$. Let $\mathcal{P}_k \subset \{1, \ldots, K\}$ denote the subset of the UEs assigned to the pilot t_k, including UEs k. The received pilot signal at TRP l can be expressed as [7]

$$\mathbf{y}^{\text{pilot}}_{t_k l} = \sum_{i \in \mathcal{P}_k} \sqrt{\tau_p \eta_i}\, \mathbf{h}_{il} + \mathbf{n}_{t_k l}, \tag{6}$$

where $\mathbf{n}_{t_k l} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma^2_{ul}\mathbf{I}_N)$ denotes the noise and η_i is the power that UEs i transmits in the UL direction. In the distributed implementation, the channel estimate is given by

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p \eta_k}\mathbf{R}_{kl}\boldsymbol{\Psi}^{-1}_{t_k l}\mathbf{y}^{\text{pilot}}_{t_k l}, \tag{7}$$

where $\boldsymbol{\Psi}_{t_k l} = \sum_{i \in \mathcal{P}_k} \eta_i \tau_p \mathbf{R}_{kl} + \sigma^2_{\text{ul}}\mathbf{I}_N$. In centralized implementation, $\mathbf{y}^{\text{pilot}}_{t_k l}$ is quantized and transmitted over the fronthaul to the CPU, where the channel is estimated as in (7), by using $\widetilde{\mathbf{y}}^{\text{pilot}}_{t_k l} \approx \alpha_{p,l}\mathbf{y}^{\text{pilot}}_{t_k l} + \boldsymbol{q}_{t_k}$ and $\widetilde{\boldsymbol{\Psi}}_{t_k l} \approx \alpha^2_{p,l}\boldsymbol{\Psi}_{t_k l} + \alpha_{p,l}(1 - \alpha_{p,l})\boldsymbol{\Psi}_{t_k l}$ [21]. The term $\boldsymbol{q}_{t_k} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{\boldsymbol{q}_{t_k}}) \in \mathbb{C}^{N\times 1}$ refers to the quantization noise with the covariance $\boldsymbol{R}_{\boldsymbol{q}_{t_k}} = \alpha_{p,l}(1 - \alpha_{p,l})\boldsymbol{\Psi}_{t_k l}$, and $\alpha_{p,l}$ represents a distortion factor associated with the number of bits b^p_l used for quantization, with the connection obtained from [31], [32].

B. Downlink Data Transmission

In UC systems, each UEs is associated with a subset of TRPs called a TRP cluster, represented by $\mathcal{M}_k \subset \{1, \ldots, L\}$. The connections between UEs k and the TRPs are denoted by a diagonal matrix $\mathbf{D}_{kl} \in \mathbb{N}^{N\times N}$, defined as

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & \text{if } l \in \mathcal{M}_k \\ \mathbf{0}_N & \text{if } l \notin \mathcal{M}_k. \end{cases} \tag{8}$$

TABLE I
COMBINING VECTORS FOR DIFFERENT PRECODING SCHEMES [7]

Scheme	Combining vector
P-MMSE ($\mathbf{v}_k$)	$\eta_k \left(\sum_{i\in\mathcal{S}_k} \eta_i \mathbf{D}_k \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\text{H}} \mathbf{D}_k + \mathbf{Z}_{\mathcal{S}_k} + \sigma^2_{\text{ul}}\mathbf{I}_{LN}\right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k$
P-RZF ($\mathbf{v}_k$)	$\left[\mathbf{D}_k \widehat{\mathbf{H}}_{\mathcal{S}_k} \left(\widehat{\mathbf{H}}^{\text{H}}_{\mathcal{S}_k} \mathbf{D}_k \widehat{\mathbf{H}}_{\mathcal{S}_k} + \sigma^2_{\text{ul}} \mathbf{P}^{-1}_{\mathcal{S}_k}\right)^{-1}\right]_{:,1}$
LP-MMSE ($\mathbf{v}_{kl}$)	$\eta_k \left(\sum_{i\in\mathcal{D}_l} \eta_i \left(\widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}^{\text{H}}_{il} + \mathbf{C}_{il}\right) + \sigma^2_{\text{ul}}\mathbf{I}_N\right)^{-1} \mathbf{D}_{kl}\widehat{\mathbf{h}}_{kl}$
MR ($\mathbf{v}_{kl}$)	$\widehat{\mathbf{h}}_{kl}$

The subset of UEs served by a TRP is denoted by $\mathcal{D}_l \subset \{1, \ldots, K\}$. In this work, the number of UEs that each TRP can serve ($K_l = |\mathcal{D}_l|$) and the number of TRPs connected to each UEs ($L_k = |\mathcal{M}_k|$) are limited, such that $K_l \leq \tau_p$ and $L_k \leq C_{\max}$. This is performed in order to comply with the scalability requirements and to reduce the processing complexity of the network [5], [20], [33]. A lower bound of the DL SE considering the fronthaul quantization can be obtained as [21], [23]

$$\text{SE}_k = \frac{\tau_d}{\tau_c}\log_2\left(1 + \frac{\text{DS}_k}{\text{IS}_k + \text{QN}_k - \text{DS}_k + \sigma^2_{dl}}\right), \tag{9}$$

where $\text{DS}_k = \left|\sum_{l=1}^{L} \mathbb{E}\{\widetilde{\alpha}_{kl}\mathbf{h}^{\text{H}}_{kl}\mathbf{D}_{kl}\mathbf{w}_{kl}\}\right|^2$ denotes the desired signal, $\text{IS}_k = \sum_{i=1}^{K}\mathbb{E}\{\left|\sum_{l=1}^{L}\widetilde{\alpha}_{il}\mathbf{h}^{\text{H}}_{kl}\mathbf{D}_{il}\mathbf{w}_{il}\right|^2\}$ stands for the interference, QN_k is the quantization noise, and σ^2_{dl} is the receiver noise variance. The term $\widetilde{\alpha}_{il}$ denotes the distortion factor associated with the number of bits b^d_{kl} used to quantize the DL signals, and $\mathbf{w}_{kl} \in \mathbb{C}^{N\times 1}$ denotes the precoding vector. Further, the relationship between the distortion factor and the number quantization of bits is obtained from [31], [32]. Both $\widetilde{\alpha}_{il}$ and QN_k are computed differently in each network implementation. Thus, $\widetilde{\alpha}_{il} = \alpha_{il}$ and $\text{QN}_k = \mathbb{E}\{|\sum_{l=1}^{L}\mathbf{h}^{\text{H}}_{kl}\sum_{i=1}^{K}\mathbf{D}_{il}\mathbf{w}_{il}q_{il}|^2\}$ for the distributed implementation, whereas $\widetilde{\alpha}_{il} = \alpha_l$ and $\text{QN}_k = \mathbb{E}\{|\sum_{l=1}^{L}\mathbf{h}^{\text{H}}_{kl}\mathbf{D}_{kl}\boldsymbol{q}_l|^2\}$ for the centralized implementation, where $q_{il} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2_{q_{il}})$ and $\boldsymbol{q}_l \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{\boldsymbol{q}_l}) \in \mathbb{C}^{N\times 1}$ are the quantization noises applied to the DL signals, with covariances $\sigma^2_{q_{il}} = \alpha_{il}(1 - \alpha_{il})$ and $\boldsymbol{R}_{\boldsymbol{q}_l} = \alpha_l(1 - \alpha_l)\sum_{i\in\mathcal{M}_k}\mathbb{E}\{\mathbf{w}_{il}\mathbf{w}^{\text{H}}_{il}\}$.

Motivated by the UL-DL duality, the DL precoding vectors can be selected based on the vectors $\mathbf{v}_k$ and $\mathbf{v}_{kl}$ of the UL combiners, such as [5], [7], [29]

$$\mathbf{w}_k = \frac{\sqrt{\rho_k}\mathbf{v}_k}{\sqrt{\mathbb{E}\{\mathbf{v}^{\text{H}}_k\mathbf{D}_k\mathbf{v}_k\}}}, \; \mathbf{w}_{kl} = \frac{\sqrt{\rho_{kl}}\mathbf{v}_{kl}}{\sqrt{\mathbb{E}\{\mathbf{v}^{\text{H}}_{kl}\mathbf{D}_{kl}\mathbf{v}_{kl}\}}}, \tag{10}$$

where $\rho_{kl} = \mathbb{E}\{\|\mathbf{w}_{kl}\|^2\}$ is the power allocated to UEs k from TRP l in the distributed implementation, whereas $\rho_k = \mathbb{E}\{\|\mathbf{w}_k\|^2\}$ is the transmit power assigned to UEs k by all its serving TRPs in the centralized one, where $\mathbf{w}_k = [\mathbf{w}^{\text{T}}_{k1}, \ldots, \mathbf{w}^{\text{T}}_{kL}]^{\text{T}} \in \mathbb{C}^{M\times 1}$ is the collective precoding vector. In addition, the DL power of each TRP is limited to ρ_l in both cases. Finally, the different combining vectors for the centralized implementation, with P-MMSE and P-RZF, and the distributed one, using LP-MMSE and MR, are given in Table I, where $\mathcal{S}_k = \{i : \mathbf{D}_k\mathbf{D}_i \neq \mathbf{0}_{LN\times LN}\}$, $\mathbf{Z}_{\mathcal{S}_k} = \sum_{i\in\mathcal{S}_k}\eta_i\mathbf{D}_k\mathbf{C}_i\mathbf{D}_k$, and

$\mathbf{C}_k$ is the correlation matrix of channel estimation error. The term $\widehat{\mathbf{H}}_{\mathcal{S}_k} \in \mathbb{C}^{LN\times|\mathcal{S}_k|}$ contains the stacked vectors $\widehat{\mathbf{h}}_i$, and $\mathbf{P}_{\mathcal{S}_k} \in \mathbb{R}^{|\mathcal{S}_k|\times|\mathcal{S}_k|}$ is a diagonal matrix containing the transmit powers η_i, both for $i \in \mathcal{S}_k$. Finally, $[\cdot]_{:,1}$ denotes the operation of only keeping the first column of its matrix argument. Unfortunately, (9) does not have a closed-form expression when using P-MMSE, P-RZF, and LP-MMSE precoding schemes. However, it is still possible to compute the SINR through Monte Carlo simulations.

C. Signal Processing Computational Complexity

In the centralized implementation, the CPUs perform channel estimation and combining/precoding computation [5]. In the distributed one, the TRPs perform these tasks locally, and the CPUs encode the DL data signals. The CC for combining and precoding-related operations is calculated as in [7], [29], taking into account the sum of the number of complex operations (multiplications and divisions) required in each coherence block.

For distributed MR precoding, the number of complex operations for each TRP l is computed as

$$C_l^{\text{MR}} = \left(N\tau_p + N^2\right)|\mathcal{D}_l|, \tag{11}$$

which represents only the LMMSE channel estimation procedure. For distributed LP-MMSE, it becomes

$$\begin{aligned} C_l^{\text{LP-MMSE}} = &\left(N\tau_p + N^2\right)|\mathcal{D}_l| + \frac{1}{2}(N^2+N)|\mathcal{D}_l| \\ &+ N^2|\mathcal{D}_l| + \frac{1}{3}(N^3-N) + N, \end{aligned} \tag{12}$$

which takes account the LMMSE channel estimation, matrix inversion, and product. It is noteworthy that the CC is a function of the number of UEs served by each TRP ($|\mathcal{D}_l|$) in the distributed implementation.

In the centralized precoding schemes, the number of complex operations depends on the number of TRPs serving each UEs ($|\mathcal{M}_k|$) and the number of UEs that are served by partially the same TRPs as UEs k, denoted as $\mathcal{S}_k = \{i : \mathbf{D}_k\mathbf{D}_i \neq \mathbf{0}_{LN\times LN}\}$. For P-MMSE precoding, there are tasks of different TRPs performed only once for each UEs. The common TRPs that serve UEs k and $i \in \mathcal{S}_k$ are denoted by the set $\mathcal{B}_k = \cup_{i\in\mathcal{S}_k}\mathcal{M}_i$. Then, the total number of complex operations is given by

$$\begin{aligned} C^{\text{P-MMSE}} = &\sum_{k=1}^{K}\Big\{(N\tau_p+N^2)|\mathcal{B}_k| + \frac{1}{2}((N|\mathcal{B}_k|)^2 + N|\mathcal{B}_k|) \\ &+ (N|\mathcal{M}_k|)^2 + \frac{1}{3}((N|\mathcal{M}_k|)^3 - N|\mathcal{M}_k|) + N|\mathcal{M}_k|\Big\}. \end{aligned} \tag{13}$$

For P-RZF, the total number of complex operations is

$$\begin{aligned} C^{\text{P-RZF}} = &\sum_{l=1}^{L}\Big\{(N\tau_p+N^2)|\mathcal{C}_l| + \frac{1}{2}(|\mathcal{C}_l|^2+|\mathcal{C}_l|)N\Big\} + \\ &\sum_{k=1}^{K}\Big\{|\mathcal{S}_k|^2 + N|\mathcal{M}_k||\mathcal{S}_k| + \frac{1}{3}(|\mathcal{S}_k|^3-|\mathcal{S}_k|) + |\mathcal{S}_k|\Big\}, \end{aligned} \tag{14}$$

where $\mathcal{C}_l = \cup_{l'\in\mathcal{M}_{k\in\mathcal{D}_l}}\mathcal{D}_{l'}$ denotes the set of common UEs that are served by TRP l and $l' \in \mathcal{M}_{k\in\mathcal{D}_l}$. The subset $\mathcal{C}_l$ is defined because there are common tasks regarding different UEs that are performed only once for each TRP l, such as the LMMSE channel estimation.

In terms of hardware computational complexity, the number of complex operations can be converted into giga operations per second (GOPS) required by the TRPs and the CPU. This leads to the multiplicative factor $f = 8N_{\text{DFT}}(\tau_c T_s)^{-1}\times 10^{-9}$ for the conversion from the number of complex operations into GOPS. The number of GOPS required by the TRPs is given by [22]

$$C_l = If\{C_l^{\text{est/com}} + C_l^{\text{cal/pre}}\} + C_l^{\text{DFT}} + C_l^{\text{BBF}}, \tag{15}$$

where $I \in \{0,1\}$ is a binary variable that indicates if centralized ($I=0$) or distributed ($I=1$) processing is used. In the first term, $C_l^{\text{est/com}} \in \{C_l^{\text{MR}}, C_l^{\text{LP-MMSE}}\}$ because only distributed processing is performed in the TRPs, and $C_l^{\text{cal/pre}} = N|\mathcal{D}_l|(1+\tau_d)$ represents the number of complex multiplications required to perform reciprocity calibration and precoding application. The terms $C_l^{\text{DFT}} = 8\,NN_{\text{DFT}}\log_2(N_{\text{DFT}})/T_s 10^9$ and $C_l^{\text{BBF}} = 40\,Nf_s/10^9$ are the required GOPS for the discrete Fourier transform (DFT) and the baseband filter [34], [35], respectively.

The required GOPS for the CPU is computed as [22]

$$C_{\text{CPU}} = (1-I)f\{C^{\text{est/com}} + C^{\text{cal/pre}}\} + C_{\text{CPU}}^{\text{other}}, \tag{16}$$

where $C^{\text{est/com}} \in \{C^{\text{P-MMSE}}, C^{\text{P-RZF}}\}$ because only centralized precoding is performed at the CPU, $C^{\text{cal/pre}} = \sum_{l=1}^{L} N|\mathcal{D}_l|(1+\tau_d)$ takes into account reciprocity calibration and precoding application, and $C_{\text{CPU}}^{\text{other}}$ denotes the CC in GOPS that the CPU demands to perform other operations that are common for centralized and distributed processing, given by [36]

$$\begin{aligned} C_{\text{CPU}}^{\text{other}} = &(B/B_{\text{b}})(\overline{\text{SE}}/\text{SE}_{\text{b}})C_{\text{HLN}} + (LN)^{0.5}K^{0.2}C_{\text{HLC}} \\ &+ (B/B_{\text{b}})(\overline{\text{SE}}/\text{SE}_{\text{b}})KC_{\text{ChC}} \\ &+ (B/B_{\text{b}})(\overline{\text{SE}}/\text{SE}_{\text{b}})^{1.5}KC_{\text{MD}} + (B/B_{\text{b}})K\,C_{\text{OFDM}}, \end{aligned} \tag{17}$$

where C_{HLN}, C_{HLC}, C_{ChC}, C_{MD}, C_{OFDM} are the base GOPS for the higher-layer network (HLN), higher-layer control (HLC), channel coding, layer mapping and demapping, and orthogonal frequency-division multiplexing (OFDM) modulation and demodulation per subcarrier, respectively. Moreover, B_{b} and SE_{b} represent the bandwidth and SE of the reference GOPS value. In contrast, B and $\overline{\text{SE}}$ are the adopted bandwidth and the achievable average SE.

III. Power Consumption and Energy Efficiency

The total power consumption can be formulated as

$$P_{\text{tot}} = \sum_{l=1}^{L}\{P_l + P_{\text{fh},l} + A_l P_{\text{UAV},l}\} + P_{\text{CPU}}, \tag{18}$$

where P_l is the power consumed by each TRP for data transmission, $P_{\text{fh},l}$ is the power that the fronthaul link connecting the CPU and TRP l consumes, and P_{CPU} is the power required by the

CPU to perform the signal processing tasks. Additionally, $P_{\text{UAV},l}$ represents the power required for a UAV to fly to a location and hover at a fixed point, and $A_l \in \{0,1\}$ is a binary variable that indicates if TRP l is terrestrial ($A_l = 0$) or aerial ($A_l = 1$).

The total EE in bit/Joule is computed as the ratio between the sum throughput in bit/s and the total power consumed in watts (W) [37], [38]. It can be written as

$$\text{EE}_{\text{tot}} = \frac{B \sum_{k=1}^{K} \text{SE}_k}{P_{\text{tot}}}, \tag{19}$$

where B is the bandwidth.

A. Power Consumption of Each TRP

The power consumed by each UAV, including data transmission and the effects of the amplifier and the circuit of the analog front-end, is modeled as

$$P_l = \frac{1}{\gamma_l} \mathbb{E}\left\{ \|\mathbf{x}_l\|^2 \right\} + N P_{\text{tc},l} + P_{\text{CC},l}, \tag{20}$$

where $0 < \gamma_l \leq 1$ denotes the efficiency of the power amplifier, and $P_{\text{tc},l}$ is the power required of each antenna of TRP l to run internal components, such as converters and filters, and $P_{\text{CC},l}$ is the power required to perform the signal processing tasks at the TRPs, given by

$$P_{\text{CC},l} = P_{l,0}^{\text{proc}} + \Delta_l^{\text{proc}} \left(\frac{C_l}{C_l^{\max}} \right), \tag{21}$$

where $P_{l,0}^{\text{proc}}$ denotes the power consumed by each digital signal processor (DSP) of TRP l in idle mode, Δ_l^{proc} represents the slope of power consumption due to processing, and $C_l^{\max}$ indicates the maximum GOPS capacity of the DSP.

B. UAV Propulsion Power Consumption

It is assumed that each UAV flies to the target deployment point, hovers during the service time, and returns to the initial point. Depending on the type of flight, the propulsion power of the UAV is $P_{\text{UAV},l} = P_{\text{hv},l}$ if it hovers or $P_{\text{UAV},l} = P_{\text{cv},l}$ if the UAV is flying.

The power required for a UAV to hover is given by [39]

$$P_{\text{hv},l} = \frac{(m_{\text{tot}} g)^{3/2}}{\eta \sqrt{2 r \rho \varsigma}}, \tag{22}$$

where m_{tot} is the total mass of each UAV, including the battery and TRP components, g is the gravity acceleration, ρ is the air density, and η denotes the battery and motor power transfer efficiency. Moreover, r represents the number of rotors of a rotocopter drone, and ς stands for the area of the spinning blade disc of one rotor.

The power required for UAV l to fly at a constant air velocity v_l in m/s is given by [40]

$$P_{\text{cv},l} = \frac{T\,(v_l \sin\alpha + v_{\text{ind},l})}{\eta}, \tag{23}$$

where $T = W + D = m_{\text{tot}} g + \frac{1}{2}\rho\, C_D\, A\, v_l^2$, and $\alpha = \tan^{-1}(\frac{D}{W})$. C_D and A are the drag coefficient of the drone and the projected area, respectively. The induced velocity $v_{\text{ind},l}$ can be numerically computed by

$$v_{\text{ind},l} = \frac{m_{\text{tot}} g}{2 r \rho \varsigma \sqrt{(v_l \cos\alpha)^2 + (v_l \sin\alpha + v_{\text{ind},l})^2}}. \tag{24}$$

C. Fronthaul Traffic Rate and Power Consumption

The rate in the fronthaul links also varies depending on the type of network implementation, i.e., centralized and distributed processing. For distributed processing, the rate scales with the number of UEs served by the TRP, whereas in the centralized implementation, it scales with the number of antennas in the TRP. The fronthaul traffic is modeled as [21]

$$R_{\text{fh},l} = 2B \left[I \frac{\tau_d}{\tau_c} \sum_{k \in \mathcal{D}_l} b_{kl}^d + (1 - I) N \left(b_l^d \frac{\tau_d}{\tau_c} + b_l^p \frac{\tau_p}{\tau_c} \right) \right], \tag{25}$$

where b_{kl}^d is the number of quantization bits per sample of each UEs used for data transmissions in the distributed implementation, and b_l^d and b_l^p are the quantization bits per sample used for data and pilot signals in each TRP, respectively, in the centralized implementation. It is assumed that $b_l^p = b^{\max}$, with $b^{\max}$ being the maximum number of quantization bits per sample. The purpose is to transmit pilots at the maximum resolution to reduce quantization errors in channel estimation. As in [23], the number of quantization bits per sample used for DL data transmission ($b^d \in \{b_l^d, b_{kl}^d\}$) is iteratively incremented in order to keep SE under small degradation, i.e., $\text{SE}_k(b^d = \infty) - \text{SE}_k(b^d) \leq a_{\text{deg}}\, \forall k \in \{1,\ldots,K\}$, where a_{deg} is a design parameter. In [23], the same bit width is applied at the TRP level, in such a way that $b_l^d = b^d$, $b_{kl}^d = b_k^d$, $\forall l \in \{1,\ldots,L\}$, and the fronthaul traffic is not restricted[1]. Conversely, we consider the same algorithms but with an additional fronthaul rate restriction, i.e., $R_{\text{fh},l} \leq R_{\text{fh}}^{\max}\, \forall l \in \{1,\ldots,L\}$. For this purpose, the maximum bit width is computed by replacing $R_{\text{fh}}^{\max}$ in (25), and choosing the minimum value between the two approaches.

The power that the fronthaul link consumes is modeled as

$$P_{\text{fh},l} = P_{0,l} + P_{\text{ft},l}\, R_{\text{fh},l}, \tag{26}$$

where $P_{0,l}$ is a fixed power consumption of each fronthaul (traffic-independent power), which may depend on the distances between the TRPs and the CPU and the system topology, $P_{\text{ft},l}$ is the traffic-dependent power (in watt per bit/s), and $R_{\text{fh},l}$ is the rate in each fronthaul link, which is different depending on the type of network implementation, i.e., centralized and distributed processing.

D. CPU Power Consumption

The power consumed by the CPU can be expressed as

$$P_{\text{CPU}} = P_{\text{fixed}} + \frac{1}{\sigma_{\text{cool}}} \left(N_{\text{GPP}} P_{\text{GPP},0}^{\text{proc}} + \Delta_{\text{GPP}}^{\text{proc}} \frac{C_{\text{CPU}}}{C_{\text{GPP}}^{\max}} \right), \tag{27}$$

[1]The values of b_l^p, b_l^d, and b_{kl}^d can be further optimized for each TRP. However, optimal bit width allocation is outside the scope of this paper.

where P_{fixed} represents the fixed power consumption of each CPU, $0 < \sigma_{\text{cool}} \leq 1$ denotes the cooling efficiency, $N_{\text{GPP}} = \lceil C_{\text{CPU}}/C_{\text{GPP}}^{\max} \rceil$ is the number of active general purpose processors (GPPs), $P_{\text{GPP},0}^{\text{proc}}$ is the idle power consumption, $\Delta_{\text{GPP}}^{\text{proc}}$ stands for the slope of power consumption, and $C_{\text{GPP}}^{\max}$ is the maximum processing capacity in GOPS.

IV. UAV Trajectory Optimization

This section introduces the algorithms to model the UEs mobility and to optimize the trajectory of the UAVs. Recall that in the context of UC D-mMIMO networks, UAVs act as mobile TRPs, and thus, their positioning in the coverage area is intrinsically associated with the system performance. Therefore, the decision process to update the positions of UAVs can be based on network performance metrics, such as SE and EE, as computed in (9) and (19). However, such metrics can change frequently in wireless channels due to channel fading and UEs mobility.

An alternative solution involves using a discretization framework, where the positioning of UAVs can be adjusted in each time frame, in conjunction with the large-scale fading coefficients, which remain unchanged for many coherency blocks. This consideration allows the calculation of a tractable ergodic SE of UEs using only large-scale channel state information (CSI). The SE is calculated on a given disposition of UEs and TRPs in each time frame. For the MR precoding, the closed-form expressions for SE with fronthaul quantization can be computed as [21], [23], [41]

$$\text{DS}_k = \alpha_k^2 \left| \sum_{l \in \mathcal{M}_k} \sqrt{\rho_{kl}\eta_k\tau_p \text{tr}(\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_k l}^{-1}\mathbf{R}_{kl})} \right|^2, \quad (28)$$

$$\begin{aligned}\text{IS}_k + \text{QN}_k = \sum_{i=1}^{K} \alpha_i \Bigg(& \sum_{l \in \mathcal{M}_i} \widehat{\rho}_{il}\eta_i\tau_p \text{tr}\left(\mathbf{R}_{kl}\boldsymbol{\Omega}_{il}\right) \\ & + P_{ik} \Bigg\{ \sum_{l \in \mathcal{M}_i} \widehat{\rho}_{il}\eta_i\eta_k\tau_p^2 \Big[\text{tr}\big(\widetilde{\mathbf{R}}_{kl}^2 \boldsymbol{\Psi}_{t_i l}^{-1}\boldsymbol{\Omega}_{il}\big) \\ & + 2\text{tr}(\boldsymbol{\Omega}_{il}\boldsymbol{\Psi}_{t_i l}^{-1}\mathbf{L}_{kl}\widetilde{\mathbf{R}}_{kl}) - \text{tr}\big((\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_i l}^{-1}\mathbf{R}_{il})^2\big) \Big] \\ & + \left| \sum_{l \in \mathcal{M}_i} \sqrt{\widehat{\rho}_{il}\eta_i\eta_k\tau_p^2} \text{tr}(\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_i l}^{-1}\mathbf{R}_{il}) \right|^2 \Bigg\} \Bigg), \end{aligned} \quad (29)$$

where α_k is the distortion factor associated with the number of bits b_k^d used for data quantization, $\boldsymbol{\Omega}_{il} = \mathbf{R}_{il}\boldsymbol{\Psi}_{t_i l}^{-1}\mathbf{R}_{il}$, $\widehat{\rho}_{il} = \rho_{il}/\text{tr}(\eta_i\tau_p\boldsymbol{\Omega}_{il})$, $\mathbf{L}_k = \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\text{H}}$, and $P_{ik} \in \{0, 1\}$ is a binary variable that indicates if different UEs share the same pilot sequence, i.e., $P_{ik} = 1$ if $i \in \mathcal{P}_k$, and $P_{ik} = 0$ if $i \notin \mathcal{P}_k$. Moreover, (28) represents the desired signal power, and (29) is the interference signals power. One can observe that using more complex precoding schemes with superior interference cancellation capabilities, such as P-MMSE and LP-MMSE, do not allow SE computation based on large-scale gains, as there are no closed-form equations available for non-MR linear precoding. Despite this limitation, the network can be optimized to MR while operating with any type of precoding. Although this strategy does not yield a globally optimized performance, it balances optimization complexity and overall system performance.

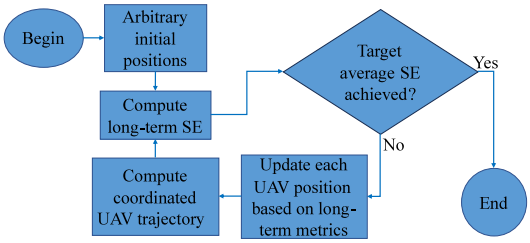


Fig. 2. Flowchart outlining the process for determining whether or not to update the positions of UAVs.

The flowchart depicted in Fig. 2 provides an overview of the approach employed in this paper to update the positioning of UAVs. Initially, all UAVs are placed in arbitrary positions, e.g., randomly or evenly spaced, and then, ergodic metrics based on long-term statistics are evaluated. If a target SE is not achieved, the new candidate positions of the UAVs are calculated based on long-term metrics, and the positions that lead to the highest performance metric are selected, as detailed in SubSection IV-B. Then, the coordinated trajectories of UAVs are computed in order to avoid collisions when UAVs move. The UAV positions will be updated until a target SE is achieved, otherwise, their positions will be fixed.

In this paper, n_s setups account for different UEs and UAVs initial positions. Over n_e episodes, UEs positions change according to their mobility behaviors, while the UAV positions may either remain static or be dynamically optimized by the proposed algorithm. If optimization occurs, each episode is subdivided into n_i iterations. During these iterations, UAVs may adjust their positions in search of the optimal location iteratively, whereas UEs maintain their positions. At the end of all iterations within an episode, UAVs assume positions where the maximum SE performance was obtained to have the best initial position in the next episode. This process is then replicated across all setups to compile the statistical numerical results of this paper. The following subsections detail the proposed models for the UEs mobility and position adjustment of the UAVs, respectively.

A. Modeling of UEs Mobility

It is considered that UEs are deployed randomly in the streets of an urban scenario with multiple buildings. The UEs may move in different directions as long as they remain inside the coverage area and streets. For simplification, it is assumed that each UEs can move at angles, given in degrees, denoted by $\phi_k \in \mathcal{A}_\phi$, where $\mathcal{A}_\phi = \{\phi^{\min}, \phi^{\min} + \phi^{\text{step}}, \ldots, \phi^{\max}\} = \{a_1, a_2+, \ldots, a_{|\mathcal{A}_\phi|}\}$. Moreover, it is important to note that $\phi^{\min} \geq 0$ is the minimum direction angle, $0 \leq \phi^{\text{step}} \leq 360$ is the step between the angles under consideration, and $\phi^{\max} \leq 360$ is the maximum direction angle.

A discrete-time Markov chain with $|\mathcal{A}_\phi| + 1$ states is used to model the mobility of each UEs, as shown in Fig. 3. There

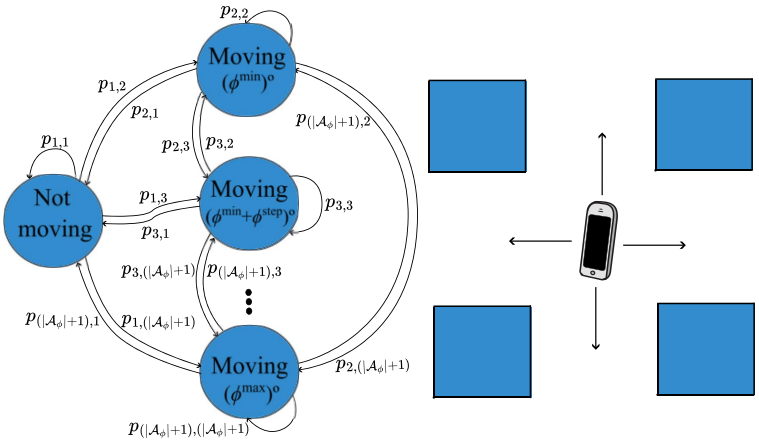


Fig. 3. Moving directions of UEs modeled by a discrete-time Markov chain, where current movement directions depend on the previous ones. The UEs positions are a function of their speeds and current moving directions.

is an additional state concerning the number of possible movement directions to model the stationary positioning behavior. In this model, a probability transition matrix $\mathbf{P} = [p_{i,j}]$, with $i \in \{1, \ldots, |\mathcal{A}_\phi| + 1\}$ and $j \in \{1, \ldots, |\mathcal{A}_\phi| + 1\}$ is established. In this context, $p_{1,1}$ is the probability of a stationary UEs to continue in this condition, which is given by p_{s}. Conversely, $p_{1,i}$ for all $i \in \{2, \ldots, |\phi_k| + 1\}$ represents the probability of a UEs initiating movement in any direction, being calculated as $(1 - p_s)/|\mathcal{A}_\phi|$, implying an equal probability of moving in any direction. Furthermore, $p_{i,1}$ for all $i \in \{2, \ldots, |\mathcal{A}_\phi| + 1\}$ indicates the probability of a moving UEs coming to a halt, which is given by p_h.

The probabilities $p_{i,j}$ for $i \neq 1$ and $j \neq 1$ are hard to calculate. UEs in motion are expected to have specific destinations, leading to a generally consistent direction of movement with only minor angle adjustments, except for sudden and more significant changes at certain locations, such as intersections or possible backtracking. These considerations imply that the probability of UEs maintaining a similar direction of movement is very high, while the likelihood of adopting an opposite direction is very low. To model this behavior, the movement direction of a UEs in an episode n, where $n > 1$, if it does not come to a halt, is given by a random variable $\phi_k^n \sim \mathcal{F}(\phi_k^{n-1}, \sigma_\phi, 180)\%360$, where $\%$ stands for the modulus operator, $\mathcal{F}(x, \sigma^2, 180) \sim \mathcal{TN}(x, \sigma^2, x - 180, x + 180)$ stands for a truncated normal distribution function, σ_ϕ is a variance that models the commitment of the UEs to maintain a similar direction of movement. Under this distribution, the probability $p_{i,j}$ for $i \neq 1$ and $j \neq 1$ is given by

$$p_{i,j} = (1 - p_h) P\left(\phi_k^n = a_j | \phi_k^{n-1} = a_i\right). \tag{30}$$

Finally, the UEs positions in the two-dimensional Cartesian space in the episode n are modeled as

$$\begin{aligned} x_k^n &= v_k^n \Delta_n \cos\left(\pi \phi_k^n / 180\right) + x_k^{n-1} \\ y_k^n &= v_k^n \Delta_n \sin\left(\pi \phi_k^n / 180\right) + y_k^{n-1}, \end{aligned} \tag{31}$$

where Δ_n is the time step between episodes. It is noteworthy that the boundaries of the buildings and the coverage area serve as obstacles. Hence, whenever a UEs approaches them, it can only move away in the opposite direction. Moreover, in each episode, the UEs move at different speeds up to a maximum value $v_{\max}$, which is randomly defined as $v_k^n \sim \mathcal{U}[0, v_{\max}]$ for each UEs k.

B. 3D Trajectory of UAVs

This subsection presents the steps to adjust the UAV positions. An iterative method is designed with the goal of improving the SE performance of the UEs. The algorithm works by defining the total number of iterations it has available to operate while the UEs are assumed to be in a stationary position. For that purpose, the number of iterations is defined as the ratio between the maximum velocities of the UAVs and the UEs, i.e., $\mathrm{n_i} = \max_{l \in \{1,\ldots,L\}}(v_l) / \max_{k \in \{1,\ldots,K\}}(v_k)$, where v_l and v_k are the velocities of UAV l and UEs k, respectively. The impact of the UAV positions on the performance of the UEs can be computed based on long-term CSI. Specifically, for any combination of UEs and UAV locations, the SE of the UEs can be obtained by replacing (28) and (29) in (9). Although these equations are specific for MR precoding, they are still useful to provide a good estimate of the UEs performance even if another precoding technique is later performed.

The first step of the algorithm is to define a set of UAVs chosen to update their positions in the current episode. This is performed based on the average SE obtained at the start of a new episode. Specifically, the set $\mathcal{M}_{k^{\min}} \subset \{1, \ldots, L\}$ defines the UAVs that serve the UEs with the worst SE performance, where $k^{\min} = \arg\min_k \mathrm{SE}_k$. It is important to note that only a subset of UAVs update their positions in each episode due to the limited number of iterations ($\mathrm{n_i}$) available.

In the next step of the algorithm, the total number of iterations is divided by the number of UAVs in the set ($\mathrm{n_i}^{\mathrm{UAV}} = \mathrm{n_i} / |\mathcal{M}_{k^{\min}}|$). The goal of this step is that each UAV adjusts its position individually, guaranteeing that the SE is not affected by

the movements of other UAVs. At each iteration, the UAV defines its moving direction $\delta_l^i = \{0, \delta^{\text{step}}, \ldots, 360\}$ (in degrees), and the UAV l position in the two-dimensional Cartesian space is modeled as

$$x_l^i = v_l \Delta_i \cos(\pi \delta_l^i / 180) + x_l^{i-1}$$
$$y_l^i = v_l \Delta_i \sin(\pi \delta_l^i / 180) + y_l^{i-1}, \tag{32}$$

where Δ_i is the time step between iterations.

The UAVs included in the set $\mathcal{M}_{k^{\min}}$ take turns adjusting their positions one at a time, where each one has $\mathrm{n_i}^{\text{UAV}}$ iterations available. The selected UAV begins by moving in a random direction, drawn from a uniform discrete distribution $\mathcal{U}[0, 360]$. Subsequently, it computes the SE improvement or reduction related to this movement, written as

$$\mathrm{SE}_{\text{dif}}^i = \frac{1}{K} \sum_{k=1}^{K} \left(\mathrm{SE}_k^i - \mathrm{SE}_k^{i-1} \right). \tag{33}$$

In the next iteration, the movement direction of the UAV is determined based on the result of the previous action. If the previous direction led to an increase in the SE, the UAV is likely to move in a similar direction. Conversely, if the previous direction resulted in a decrease in the SE, the UAV is more likely to move in the opposite direction. In both cases, the next direction is chosen randomly from a normal distribution with a mean equal to the previous direction (or its opposite) and a variance calculated based on the SE improvement (or decrease) of the previous action. However, if there is no SE improvement or decrease, the UAV moves in a uniformly random direction. Therefore, the direction of UAV l movement at iteration i is modeled as

$$\delta_l^i \sim \begin{cases} \mathcal{U}[0, 360] \text{ if } \mathrm{SE}_{\text{dif}}^{i-1} = 0 \text{ or } i = 1, \\ \mathcal{F}(\delta_l^{i-1}, \sigma_\delta^{i-1}, 180) \% 360 \text{ if } \mathrm{SE}_{\text{dif}}^{i-1} > 0, \\ \mathcal{F}(\bar{\delta}_l^{i-1}, \sigma_\delta^{i-1}, 180) \% 360 \text{ if } \mathrm{SE}_{\text{dif}}^{i-1} < 0, \end{cases} \tag{34}$$

where $\%$ stands for the modulus operator, $\bar{\delta}_l^{i-1} = (180 + \delta_l^{i-1}) \% 360$ is the opposite angle of δ_l^{i-1}, $\mathcal{F}(x, \sigma^2, 180) \sim \mathcal{TN}(x, \sigma^2, x - 180, x + 180)$ is a truncated normal distribution function with the mean value centered at x degrees and variance σ^2. The variance parameter is computed based on the SE improvement (or reduction) as

$$\sigma_\delta^{i-1} = \min\left(\sigma_\delta^{\max}, \frac{1}{\left| \mathrm{SE}_{\text{dif}}^{i-1} \right|} \right), \tag{35}$$

where $\sigma_\delta^{\max}$ is a project parameter. At the end of the iterations, UAV l adjusts its position to the iteration that led to the highest SE performance, i.e., $i^\star = \arg\max_i \sum_{k=1}^{K} \mathrm{SE}_k^i$. The algorithm concludes once all the UAVs adjust their positions. Algorithm 1 summarizes the heuristic method proposed to adjust the UAV positions.

Furthermore, the heights z_l^i of UAVs can also be dynamically optimized in each iteration, where $z_{\min} \leq z_l^i \leq z_{\max}$. In this case, a fraction of the iterations allocated for each selected UAV is used to adjust its height while its two-dimensional position remains the same, i.e., $[x_l^i = x_l^{i-1},\ y_l^i = y_l^{i-1}]$. For that purpose, a modified version of Algorithm 1 is used, considering that each UAV moves upward or downward in a given iteration, represented as $z_{l,(\mathrm{u/d})}^i \in \{-1, 1\}$. Additionally, the UAV l height is modeled as $z_l^i = v_l\, \Delta_i\, z_{l,(\mathrm{u/d})}^i + z_l^{i-1}$. Afterward, if the previous action resulted in a SE increase, it continues to move in that direction. Otherwise, it goes in the opposite direction. The process may stop before the end of the assigned iterations if the UAV reaches the minimum or maximum heights.

Algorithm 1: Iterative Adjustment Method for the Positions of Selected UAVs per Episode.

Input: Current episode n, number of iterations $\mathrm{n_i}$, maximum variance parameter $\sigma_\delta^{\max}$, $\mathrm{SE}_{\text{dif}}^{(1)} = 0$.

`// Compute SE at start of episode n`

Compute $\mathrm{SE}_k^{(i=0)} = \mathrm{SE}_k^n \ \forall k \in \{1, \ldots K\}$ by solving (28) and (29) in (9);

`// Select UAVs serving the worst UE`

Set $k^{\min} = \arg\min_k \mathrm{SE}_k^n$;

Set $\mathcal{M}_{k^{\min}} \subset \{1, \ldots, L\}$ as the subset of UAVs that serve the UE with the worst SE performance;

`// Define iterations per UAV`

Set $\mathrm{n_i}^{\text{UAV}} = \mathrm{n_i} / |\mathcal{M}_{k^{\min}}|$;

for $l = 1$ **to** $|\mathcal{M}_{k^{\min}}|$ **do**

  for $i = 1$ **to** $\mathrm{n_i}^{\text{UAV}}$ **do**

    if $i = 1$ **or** $\mathrm{SE}_{\text{dif}}^{\mathrm{i}} = 0$ **then**

      `// Move in a random direction`

      Set $\delta_l^i \sim \mathcal{U}[0, 360]$;

    end

    else if $\mathrm{SE}_{\text{dif}}^{\mathrm{i}} < 0$ **then**

      `// Move in opposite direction`

      Set opposite angle as $\bar{\delta}_l^{i-1} = (180 + \delta_l^{i-1}) \% 360$;

      Set $\delta_l^i \sim \left[\mathcal{F}(\bar{\delta}_l^{i-1}, \sigma_\delta^{i-1}, 180) \right] \% 360$;

    end

    else

      `// Move in similar direction`

      Set $\delta_l^i \sim \left[\mathcal{F}(\delta_l^{i-1}, \sigma_\delta^{i-1}, 180) \right] \% 360$;

    end

    Set $\begin{matrix} x_l^i = v_l \Delta_i \cos(\pi \delta_l^i / 180) + x_l^{i-1} \\ y_l^i = v_l \Delta_i \sin(\pi \delta_l^i / 180) + y_l^{i-1} \end{matrix}$;

    Compute $\mathrm{SE}_k^i \ \forall k \in \{1, \ldots K\}$ by solving (28) and (29) in (9);

    Set $\mathrm{SE}_{\text{dif}}^i = \frac{1}{K} \sum_{k=1}^{K} (\mathrm{SE}_k^i - \mathrm{SE}_k^{i-1})$;

    Set $\sigma_\delta^i = \min\left(\sigma_\delta^{\max}, \frac{1}{|\mathrm{SE}_{\text{dif}}^i|} \right)$;

  end

  `// UAV position at highest SE`

  Set $i^\star = \arg\max_i \sum_{k=1}^{K} \mathrm{SE}_k^i$;

  Set $[x_l, y_l] = [x_l^{i^\star}, y_l^{i^\star}]$ as the value of iteration $i^\star$;

end

Output: Updated positions: $[x_l, y_l] \ \forall l \in \mathcal{M}_{k^{\min}}$

Note that only channel statistics are needed to compute the proposed UAV trajectory optimization in Algorithm 1, and that it considers imperfect CSI to calculate the the SE of the UEs. Analyzing the proposed algorithm, the time complexity can be written as $\mathcal{O}(\mathrm{n_i} K^2 |\mathcal{M}_i| N^3)$, which is due to the computation of

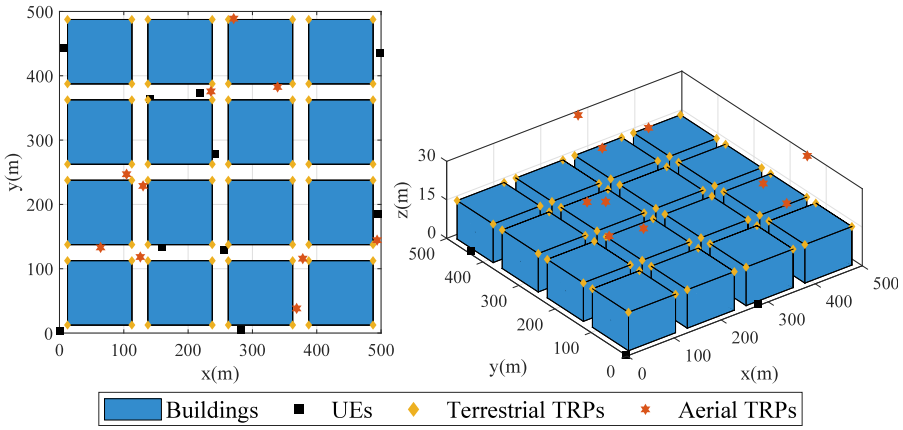


Fig. 4. Urban-dense scenario under consideration to analyze UC D-mMIMO network enabled by a swarm of UAVs as TRPs. In scenarios containing terrestrial TRPs, they are placed atop the building's edges.

the SE of the UEs for each iteration. For scalability purposes, this work assumes that $|\mathcal{M}_i| \leq C_{\max}$, and that the CPU can compute the SE of only $K \leq K_{\max} = LC_{\max}$ UEs, corresponding to the number of connections in the network [33]. Then, the time complexity of the algorithm can be defined as $\mathcal{O}(\mathrm{n_i} K_s^2 C_{\max} N^3)$, where $K_s = \min(K, K_{\max})$.

V. Numerical Results

This section provides illustrative numerical results of the UC D-mMIMO system enabled by UAVs. The results for the proposed UAV trajectory optimization are presented for scenarios where only aerial TRPs are deployed and also for the case where both aerial and terrestrial TRPs are used. The numerical results focus on the achievable DL SE, which serves as a useful bound and provides sufficient insights regarding the efficiency of the proposed trajectory optimization in UAV-enabled UC D-mMIMO systems.

The simulation scenario has a 500 m x 500 m square coverage area, with 16 building blocks of 100 m square length and 15 m high, as depicted in Fig. 4. The grid structure is common in major cities such as Barcelona and New York, making it a suitable model for urban-dense environments. The UEs are randomly deployed throughout the streets of the scenario. For scenarios containing terrestrial TRPs, they are located at the top of building edges. Initially, the aerial TRPs are carried by UAVs hovering above the streets at a height of 30 m. The positioning and heights of the UAVs can be changed by performing the proposed trajectory optimization method in Algorithm 1, where the UAVs always fly at heights above the buildings, and at the height of 100 m at the maximum. To determine the presence of a LoS link between each UEs and TRP, we check whether the buildings in Fig. 4 physically obstruct the LoS component or not. Table II presents the remaining simulation parameters.

The maximum number of UEs is based on a high-density urban area with 10,000 people per km^2, 25% of the UEs being outdoor UEs, and the operator has a contract with one-third of the outdoor UEs. Three UEs density scenarios are simulated to emulate different load behaviors. The corresponding number of UEs according to the coverage area is shown in Table III. The figures given in the table represent low, medium, and high UEs densities, which are equivalent to 144, 520, and 800 UEs/km^2, respectively. The number of deployed TRPs depends on the number of UEs, antennas per TRP, and the expected data rate of 200 Mbit/s demand per UEs using MR precoding. The procedure to determine the number of deployed TRPs is resumed by computing the average rate R for a given value of L, N, and K, using Monte Carlo simulations. The set of TRP count and the achieved rate are stored in sets $\mathcal{L} = \{L_{\min}, L_{\min} + L_{\text{step}}, \ldots, L_{\max}\}$ and $\mathcal{R} = \{R_1, R_2, \ldots, R_{|\mathcal{L}|}\}$, respectively. By noting that $R_1 < R_2 < \cdots < R_{|\mathcal{L}|}$, it is possible to interpolate the results to determine any arbitrary L that leads to an average rate R. To obtain the values in Table III, $L_{\min} = K/\tau_p$ and $L_{\max} = 800$. Because the results are depicted in terms of the SE throughout this paper, note that 200 Mbit/s per UEs is equivalent to 2 bit/s/Hz with the adoption of a 100 MHz bandwidth.

TABLE II
System, Channel, and Signal Simulation Parameters

Parameter	Value
Heights of UEs	1.65 m
Heights of aerial and terrestrial TRPs	30 m and 16.65 m
Receiver noise figure	7 dB
Coherence interval	$\tau_c = 200$ samples
UL pilot length	$\tau_p = 10$ samples
Maximum TRPs per UE	$C_{\max} = 10$
Carrier frequency	$f_c = 3.5$ GHz
Bandwidth	$B = 100$ MHz
Total DL power of the TRP	$\rho_l = 23$ dBm
Total UL power of the UE	$\eta_k = 20$ dBm
Angular standard deviations	$\sigma_\varphi = \sigma_\theta = 15°$
ULA antenna spacing	1/2 wavelength distance
Urban environment settings	$a = 9.61$, $b = 0.16$, $\eta^{\mathrm{LoS}} = 1$, $\eta^{\mathrm{NLoS}} = 20$
Continue stationary and come to a halt probabilities	$p_s = 0.15$, $p_h = 0.3$
Time between episodes and iterations	$\Delta_n = 5/v_{\max}$, $\Delta_i = \Delta_n/\mathrm{n_i}$
Variances for UEs and UAVs	$\sigma_\phi = 5$, $\sigma_\delta^{\max} = 30$
Number of setups, episodes, and iterations	$\mathrm{n_s} = 100$, $\mathrm{n_e} = 100$, and $\mathrm{n_i} = 100$

TABLE III
NUMBER OF DEPLOYED TRPS FOR DIFFERENT UES DENSITY SCENARIOS

UE density	Number of UEs (K)	Number deployed of TRPs (L)	
		N = 2	N = 4
Low	36	84	48
Medium	130	303	167
High	200	491	265

TABLE IV
PARAMETERS FOR CALCULATING THE POWER CONSUMPTION AND EE

Parameter	Value	Parameter	Value
$P_{l,0}^{\text{proc}}$, $P_{\text{GPP},0}^{\text{proc}}$	7.3 W, 212.4 W	$P_{0,l}$	0.825 W
Δ_{l}^{proc}, $\Delta_{\text{GPP}}^{\text{proc}}$	13.64 W, 452.08 W	$P_{\text{ft},l}$	0.25 W/(Gbit/s)
$C_{l}^{\max}$	180 GOPS	σ_{cool}, γ_l	0.9, 0.4
$C_{\text{GPP}}^{\max}$	10777 GOPS	P_{fixed}	450 W

A joint UL pilot assignment and TRP selection scheme presented in [7, Sec. 4.4] is used in this study. In this method, the first τ_{up} UEs are assigned mutually orthogonal pilots, and the remaining UEs are assigned to the pilot that causes the lowest pilot contamination. After that, each TRP chooses to serve τ_{up} UEs with the largest channel gain in each pilot. In the distributed implementation, the power coefficients at TRP l are determined as $\rho_{kl} = \rho_d\sqrt{\beta_{kl}}/\sum_{k'\in\mathcal{D}l}\sqrt{\beta_{k'l}}$, where ρ_d represents the maximum transmit power per TRP. The centralized approach adopts scalable fractional power control with the parameters set to $v = -0.5$ and $\kappa = 0.5$ [7]. The EE parameters related to the power consumption of the hardware of the TRPs and the fronthaul links are summarized in Table IV, which follows [22], [38]. The DSPs parameters align with Texas Instruments TMS320C6678, and those for GPPs are based on dual Intel Xeon Gold 6338 N processors. The remaining parameters used to compute the CC in GOPS are based on conventional 5 G new radio (NR) with 30 kHz of subcarrier spacing, where $N_{\text{DFT}} = 3300$, $f_s = 122.88$ MHz, and $T_s = 35.38\,\mu$s. The parameters used to model the propulsion power are set as follows: $m_{\text{tot}} = 0.411$ kg, $\eta = 0.7$, $r = 4$, $\rho = 1.225\,\text{kg/m}^3$, $\varsigma = 0.0176\,\text{m}^2$, $C_D = 2.49$ and $A = 0.0636\,\text{m}^2$, and a battery capacity of 2.0457×10^5 J [40]. The number of bits per sample for data transmissions in the distributed and centralized implementations, b_{kl}^d and b_l^d, are computed by setting $a_{\text{deg}} = 0.1$ bit/s/Hz for the acceptable SE degradation due to fronthaul quantization, and $R_{\text{fh}}^{\max} = 10$ Gbit/s for the maximum fronthaul rate. In the centralized implementation, the bit width of the pilot samples is set as $b^{\max} = 10$.

A. Benchmark Comparison and Convergency Analyses

In Fig. 5, the proposed UAV trajectory method is compared with the benchmark particle swarm optimization (PSO) algorithm, which is used in [17]. The PSO algorithm is implemented in the MATLAB optimization toolbox using the default value for self and social adjustment weights of 1.49, an inertia weight range from 0.1 to 1.1, and a swarm size of $\text{N}_\text{p} = 100$ particles per iteration. The objective function is set to maximize the average SE, and the candidate solutions are 3D positions of the UAVs, with coordinates bounded such as $0 \le x \le 500$, $0 \le y \le 500$,

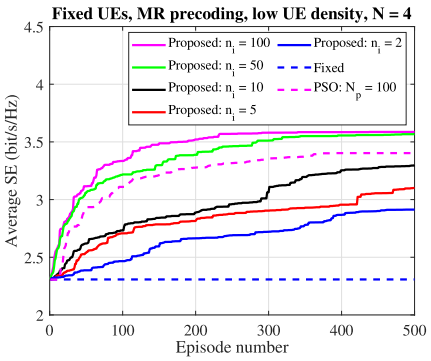


Fig. 5. Comparison of the proposed UAV trajectory optimization with the PSO algorithm. The results consider the low UE density scenario, with $N = 4$ and MR precoding, and no UE mobility.

and $15.5 \le z \le 100$, and with a minimum distance between UAVs constrained to 5 m. Note that the PSO algorithm assumes that the CPU knows the UEs positions in order to compute the objective function of all candidate solutions in each iteration. In this result, 500 episodes and PSO iterations are considered. Moreover, the UEs positions remain fixed throughout the episodes, and the same initial positions for the UEs and the UAVs are used for all curves to reduce the number of simulations and ensure a fair comparison. It can be seen that the proposed method surpasses the PSO algorithm when n_i is 50 and 100. However, with n_i equal to 2, 5, and 10, more than 500 episodes would be needed to converge. On the other hand, the PSO algorithm converges but does not find an optimal solution. This is because we have selected the default value for adjustment and inertia weights, and a fine-tuning of these parameters might be required. Nonetheless, even if the PSO converges to the optimal solution, it is essential to note that the proposed UAV trajectory method works without knowing the UEs positions, but it can also be used for UAV initial deployment in case the UEs positions are known. Further, the proposed algorithm can be used even if the UEs are moving, which is analyzed in the following subsections.

B. 3D Trajectory of UAVs

This subsection delves into cases where UAVs are deployed to provide mobile service coverage when terrestrial infrastructure is unavailable due to malfunctioning, disasters, or when it has not been deployed.

Fig. 6 depicts the average SE of UEs plotted against the episode number. This result is derived from various Monte Carlo independent simulations, with random initial positions assigned to both TRPs and UEs. The analysis assumes a MR precoding scheme for different UEs density scenarios, as outlined in Table III. The proposed algorithm for the adjustment of UAV positions is compared with the case where their positions remain fixed. It can be observed that when the UAVs are stationary, the SE tends to remain constant throughout the episode steps. This is attributed to the movement of UEs, which can yield both positive and negative impacts on the SE. Since the results are an average of different setups, one effect cancels out the

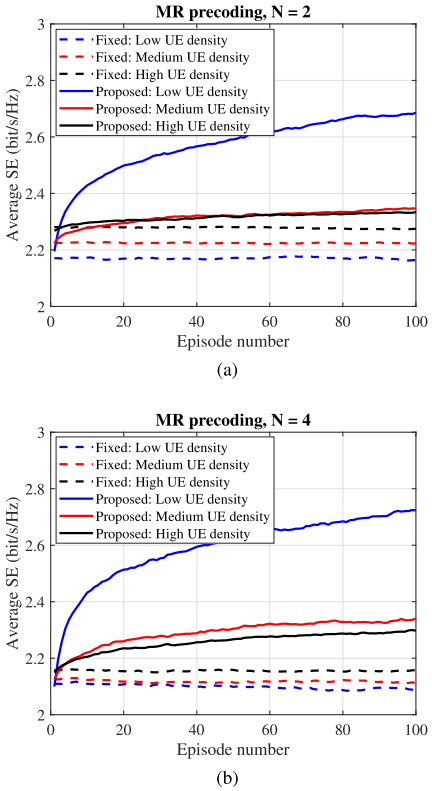


Fig. 6. Average DL SE of UEs per episode for MR precoding at low, medium, and high UE density scenarios. (a) $N = 2$. (b) $N = 4$.

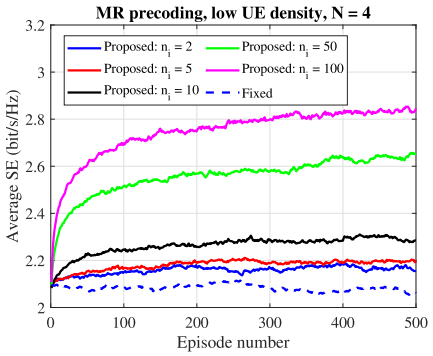


Fig. 7. Average DL SE of UEs per episode for different numbers of iterations (n_i) in the proposed UAV trajectory optimization. The results consider the low UEs density scenario, with $N = 4$ and MR precoding.

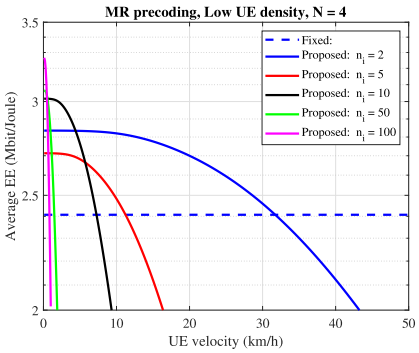


Fig. 8. Average EE vs maximum velocity of the UEs with varying numbers of iterations n_i. Results are for MR precoding, low UE density, with $N = 4$.

other, leading to a stable average SE. On the other hand, the average SE monotonically increases with the episodes when the proposed algorithm is performed. This demonstrates that UAVs can iteratively adjust their positions to improve SE within each episode, allowing for dynamic adaptation to the movement of UEs across episodes.

Moreover, the results in Fig. 6 show that the highest increase in the SE of the proposed algorithm is achieved when the UEs density is lower. For instance, with two antennas per TRP, the average SE improvement is 23.75% for a low UEs density, whereas the improvement is only 2.44% in a high UEs density. Nonetheless, the SE improvement for scenarios with four antennas per TRP is slightly better, reaching 29.76% for a low UEs density and being 6.56% in scenarios of a high UEs density. These results suggest that the proposed algorithm yields a smaller SE improvement when more UAVs are deployed, given that results for two antennas are slightly worse than for four antennas. This occurs because only a limited number of UAVs can adjust their positions within each episode. Consequently, with an increased number of deployed UAVs, the fraction of them capable of adjusting their positions decreases, leading to slightly reduced SE improvements.

Fig. 7 presents the average SE of the UEs versus the episode number for the proposed UAV trajectory optimization when the number of iterations n_i varies. The results assume MR precoding for the low UEs density scenario with four antennas per TRP, and evaluate up to 500 episodes. It is observed that the performance of the proposed method improves with an increase in the number of iterations. In episode 500, the SE improvements of the proposed approach compared with the fixed UAV positions reach 38%, 28.6%, and 10.8% for n_i equal to 100, 50, and 10, respectively. It is interesting that even with the number of iterations as low as 2 or 5, slight SE improvements are achieved, although it is clear that guaranteeing a higher number of iterations is preferable.

Nonetheless, it is important to note that the number of iterations is not always a controllable design parameter. It is computed as the velocity ratio of UAVs and UEs and represents the time frame where the UEs positions do not change considerably. Thus, if the UEs velocities are high, it may not be possible to deploy UAVs with even higher velocities. If we assume a maximum UAV velocity of 100 km/h, the allowed UEs velocities depend on the number of iterations n_i. In addition, the required propulsion power of UAVs increases with their velocities, which may reduce the EE even though the SE improves. Fig. 8 illustrates the average DL EE of the UEs vs the maximum UEs velocity with varying numbers of iterations n_i, exemplifying how the UEs and UAV velocities impact the total EE. The results in this figure assume that the average EE performance of the proposed algorithm is obtained at episode

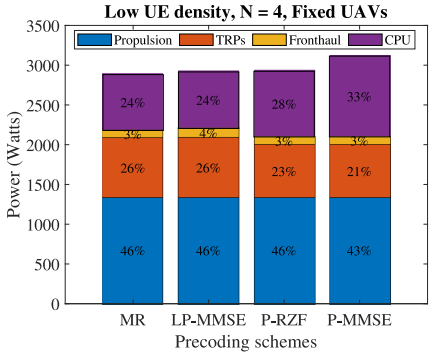


Fig. 9. Power composition in low UE density scenarios and fixed UAVs.

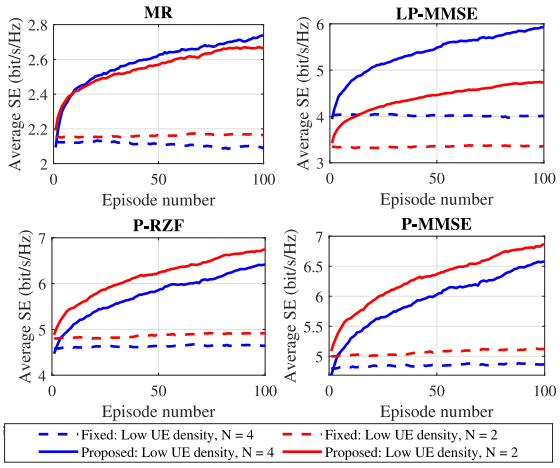


Fig. 10. Average DL SE of UEs per episode for MR, LP-MMSE, P-RZF, and P-MMSE precoding schemes. Results consider low UE density scenarios.

100. It can be noted that the EE of the proposed algorithm can be lower than the fixed positioning approach, depending on the UEs velocities and the number of iterations n_i . For instance, with a maximum UEs velocity of 20 km/h, the SE gains provided by a fivefold UAV velocity (i.e., $n_i = 5$) do not compensate in terms of the EE compared with fixed positioning.

Moreover, the intersection points where the proposed approach matches the fixed positions occur at even lower UEs speeds the higher the n_i value, as shown in Fig. 8. This illustrates the dominance of the propulsion power of UAVs over the other variables in the total power consumption, especially at higher UAV speeds. In terms of flight autonomy, the impact of the increased power consumption at high UAV speeds affects only the UAVs that are adjusting their positions. Because the proposed method alternates the selected UAVs in each episode, the reduction in the flight autonomy time of each UAV is mitigated. Essentially, each UAV moves at a higher speed for a brief period before hovering and allowing the next UAV to move. For example, if we consider a maximum speed of 10km/h for the UEs, we can determine the flight autonomy time by varying the maximum speed of the UAV. Additional results show that with a maximum UAV velocity of 20 km/h (when $n_i = 2$), the flight autonomy time is around 51min. Similarly, with a maximum UAV velocity of 100 km/h (when $n_i = 10$), the flight autonomy time reduces to 46 min. This time difference is not significant because the UAVs alternate between moving and hovering. Fig. 9 illustrates the total power consumption of UAV propulsion, data transmission and signal processing by TRPs, fronthaul traffic, and processing in the CPU. The results are shown for MR, LP-MMSE, P-RZF, and P-MMSE precoding schemes, assuming that the UAVs are hovering in fixed positions. We can see that the UAV propulsion power has the highest impact on the total power, consuming around 1333 W (27 W per UAV). Interestingly, although the maximum power required for data transmission per TRP is 200 mW, the power per TRP is also significant, which is mainly due to the DSPs. Moreover, the power consumption for data transmission over fronthaul links has the smallest impact on the total power consumption, as the fronthaul rate is limited to 10 Gbit/s, and it can be 160 W at the maximum for this scenario. Finally, considering the power in the CPU, it can be noticed that it is much higher with P-MMSE than with the other alternatives, which is expected because it is the one with the highest computational requirements.

Fig. 10 illustrates the performance of the proposed UAV trajectory optimization for MR, LP-MMSE, P-RZF, and P-MMSE precoding schemes. The results consider low UEs density scenarios. It is noteworthy that the algorithm to heuristically optimize the UAV positions can be used with precoding schemes other than MR, even though it is based on simpler MR closed-form expressions for achievable SE. Recall that the proposed algorithm uses the MR closed-form regardless of the precoding choice to take advantage of the long-term CSI. As in Fig. 6, the average SE remains constant throughout the episode steps for fixed UAV position, whereas it increases monotonically when the proposed algorithm is performed.

Furthermore, the findings depicted in Fig. 10 provide evidence of the superior performance of distributed LP-MMSE, centralized P-RZF, and P-MMSE compared with MR. For instance, their average SE can be more than twice that of MR (being approximately 2.17 to 2.41 times higher). Furthermore, the SE improvements achieved by the proposed approach are also greater, e.g., up to 47.84% for LP-MMSE, 38.14% for P-RZF, and 35.24% for P-MMSE, when compared with the fixed UAV position with four antennas per TRP. As with MR precoding, the results for the other precoding schemes also show that the four antennas per TRP configuration yields slightly higher SE improvements than with two antennas per TRP, although this is not readily observable in the figure. It is also noteworthy that LP-MMSE favors the four antennas per TRP scenario while P-RZF and P-MMSE perform better with two antennas. This is attributed to the impact of the number of antennas per TRP in their capacity to mitigate interference. Generally, centralized precoding schemes are able to combine information of different TRPs to better mitigate inference, benefiting from having more TRPs across the coverage area. On the other hand, distributed precoding uses only local information, requiring a balance in the number of TRPs and antennas per TRP.

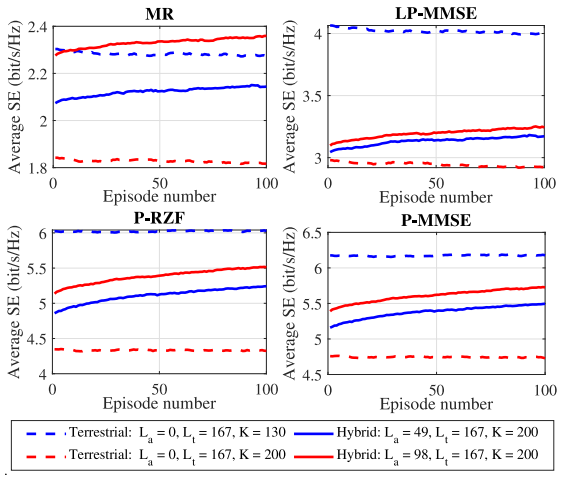


Fig. 11. Average DL SE of UEs per episode for medium and high UE density and $N = 4$. Comparison between systems consisting exclusively of terrestrial TRPs and those with a combination of terrestrial and aerial TRPs. The results are for the MR, LP-MMSE, P-RZF, and P-MMSE precoding schemes.

C. Hybrid Terrestrial and Aerial TRPs Scenario

This subsection explores scenarios where UAVs are temporarily deployed to assist in cases where the terrestrial infrastructure is still available, but a surge in demand requires additional TRPs to provide a better service experience. Specifically, hybrid scenarios that combine aerial and terrestrial TRPs are evaluated to deal with the demands of high UEs density and high UEs data rates. It is assumed that the path loss and LoS probability models for both terrestrial and aerial TRPs remain consistent, mitigating potential influences on the following result evaluation.

Fig. 11 illustrates the average DL SE of UEs per episode for medium and high UEs density and $N = 4$. The results compare D-mMIMO systems consisting exclusively of terrestrial TRPs with those with a combination of terrestrial and aerial TRPs. The figure also presents results for the MR, LP-MMSE, P-RZF, and P-MMSE precoding schemes. In the results, L_t stands for the number of terrestrial TRPs while L_a stands for the number of aerial TRPs, with the total number of TRPs defined as $L = L_t + L_a$. The simulation parameters emulate scenarios where the terrestrial infrastructure is deployed to serve a medium UEs density (i.e., $K = 130$, $L = L_t = 167$). For instance, this provides an average SE per UEs of 2.275 bit/s/Hz with MR precoding. However, when the number of UEs increases (a high UEs density with $K = 200$), the deployed terrestrial TRP can only provide 1.817 bit/s/Hz per UEs. This represents a decrease of approximately 20% due to the increased interference and competition for radio resources. Fortunately, with the aid of UAVs, it is possible to restore the per UEs performance to 2.144 bit/s/Hz if $L_a = 49$ or up to 2.359 bit/s/Hz if $L_a = 98$. In the LP-MMSE, P-RZF, and P-MMSE precoding schemes, similar behaviors can be observed. However, even after deploying additional 98 UAVs when the number of UEs is 200, the SE performance is not superior to that of the terrestrial scenario with only 130 UEs, as it happens with MR precoding. To achieve that, these precoding options would require even more additional TRPs, especially in the case of LP-MMSE. Nonetheless, these precoding choices provide superior performance to MR in any of the analyzed cases. Additionally, P-MMSE precoding is the one that achieves the highest SE performance, making it the preferable choice.

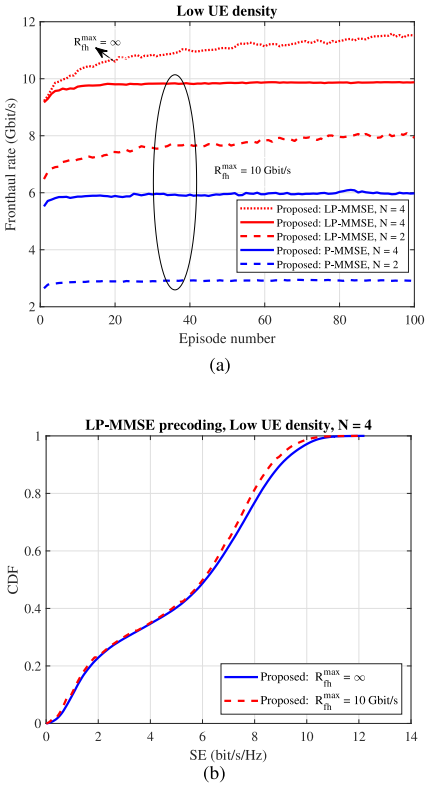


Fig. 12. (a) Required fronthaul rate per episode for the LP-MMSE and P-MMSE precoding schemes. (b) CDF of DL SE with and without limitation of the fronthaul rate. Both results are shown for low UEs density scenarios.

D. Required Fronthaul Rate

This section investigates the required rate of the wireless fronthaul links connecting the TRPs to the CPU. In Fig. 12, the results consider UC D-mMIMO systems consisting exclusively of aerial TRPs, for a low UEs density scenario, when the proposed UAV trajectory optimization is performed.

Fig. 12(a) presents the required fronthaul rate for the distributed LP-MMSE and centralized P-MMSE precoding schemes. For both precoding schemes, the required fronthaul rate increases with the number of antennas per TRP. This occurs because the achievable SE of the UEs also increases with the number of antennas per TRP. Hence, in order to maintain an acceptable SE degradation due to fronthaul data samples set as $a_{\text{deg}} = 0.1$ bit/s/Hz, the bit width also increases. We can also see that the fronthaul rate can exceed 10 Gbit/s for LP-MMSE, if it is not limited when computing bit allocation. Such behavior does not occur with P-MMSE. However, to avoid a high fronthaul traffic rate, all previous results assumed $R_{\text{fh}}^{\max} = 10$ Gbit/s for

SE, EE, and power consumption. Furthermore, it is noteworthy that the required fronthaul rate is much higher for the distributed implementation than for the centralized one. For instance, the rate can be around 2.7 times higher when comparing the distributed and centralized implementations with $N = 2$. This happens because the fronthaul rate scales with different parameters, as shown by (25). In the distributed implantation, it scales with the number of UEs served by each TRP, which is set as $|\mathcal{D}_l| \leq (\tau_p = 10)$, whereas in centralized implementations, it scales with the number of antennas per TRP N, which in the results is set either as two or four. Noting that $N \leq \tau_p$, this behavior is expected. Fig. 12(b) shows the cumulative distribution function (CDF) of DL SE with and without a maximum fronthaul rate per link constraint. The results analyze only LP-MMSE precoding because it can require more than 10 Gbit/s per link when there is no fronthaul rate limitation, as shown in Fig. 12(a). It can be seen that limiting the fronthaul rate to 10 Gbit/s has only a slight impact on the CDF of SE. Only the most fortunate UEs suffer a slight decrease in their SE performance.

VI. Conclusion

This paper investigated the performance of scalable UC D-mMIMO systems enabled by a swarm of UAVs acting as TRPs, which could operate independently or in conjunction with terrestrial TRPs. A comprehensive study on deployment and trajectory optimization of UAVs was conducted, proposing a novel heuristic approach to optimize the positions of aerial TRPs. As far as the authors are aware of, the proposed heuristic approach is the first to consider the continuous movement of UEs within the coverage area for UC D-mMIMO systems. Specifically, it optimizes the three-dimensional locations of each UAV under a time-discretized framework divided into episodes, each consisting of several iterations. The positions of UEs vary between episodes through a discrete-time Markov chain. The positions of TRPs are optimized iteratively based on the ratio between maximum velocities of UAVs and UEs, and the MR precoding closed form for SE. Moreover, each UAV adjusts its position independently, guaranteeing that the SE is not affected by the movements of other UAVs.

The evaluation of the heuristic approach addressed the achievable DL SE performance using MR, LP-MMSE, P-RZF, and P-MMSE precoding, assuming densities from 144 to 800 UEs per km^2. The results indicated that although the optimization only uses the MR equations, the method can be applied to more advanced precoding with significant performance improvements for all UEs densities under consideration. In fact, for a fully aerial network with a low UEs density, the SE enhancements were 29.76% for MR, 47.84% for LP-MMSE, 38.14% for P-RZF, and 35.24% for P-MMSE, when comparing the proposed method with the fixed UAV deployment. The most significant SE improvements were usually found at the lower UEs density and with a higher number of iterations. Despite this, in terms of the DL EE, the additional propulsion power required for UAVs to move at high speeds may not be justified by the benefits of more iterations.

Another notable insight from the results is that when the terrestrial infrastructure is overwhelmed by a surge in UEs density or demand, temporarily deploying UAVs can restore network performance in terms of the SE. Finally, the effects of fronthaul constraints were also examined. Specifically, when the fronthaul rate was limited to 10 Gbit/s, it had no significant impact on the CDF of SE compared with scenarios with unconstrained fronthaul. Additionally, when seeking a quantization level with minimal SE degradation under unrestricted fronthaul rates, the findings indicated that the distributed LP-MMSE precoding requires more fronthaul capacity than the centralized options.

Based on the observed findings and the novelty of the proposed approach, it is reasonable to suggest that this paper can serve as a benchmark for subsequent studies in this field. Future research directions might include exploring the use of fixed-wing UAVs within the proposed framework, and examining the effects of channel aging, which causes the channel estimates to become rapidly outdated due to the mobility of both UEs and UAVs. Further investigations could also assess different antenna array configurations and patterns, evaluate how weather conditions affect the UAV positioning accuracy, and consider the implications of no-fly zones that limit the operational spaces of UAVs.

Acknowledgment

This study used computational resources from the High-Performance Computer Center (https://www.ccad.ufpa.br/) at the Federal University of Pará. The authors would like to thank Dr. Hanna Niemelä for proofreading this paper.

References

[1] 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. 99878–99888, 2019.

[2] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[3] G. Interdonato, E. Björnson, H. Q. Ngo, P. K. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, Aug. 2019, Art. no. 197.

[4] 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.

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

[6] 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. 2020.

[7] Ö. Demir, E. Björnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO (Foundations and Trends in Signal Processing Series)*. Delft, The Netherlands: Now Publishers, 2021.

[8] W. Shin and M. Vaezi, *UAV-Enabled Cellular Networks*. Cham, Switzerland: Springer, 2021, pp. 165–200.

[9] "Handbook of unmanned aerial vehicles," 2015. [Online]. Available: http://dx.doi.org/10.1007/978-90-481-9707-1

[10] A. A. Khalil, M. Y. Selim, and M. A. Rahman, "CURE: Enabling RF energy harvesting using cell-free massive MIMO UAVs assisted by RIS," in *Proc. IEEE 46th Conf. Local Comput. Netw.*, Oct. 2021, pp. 533–540.

[11] O. Abbasi and H. Yanikomeroglu, "A cell-free scheme for UAV base stations with HAPS-assisted backhauling in terahertz band," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 249–254.

[12] L. Wang and Q. Zhang, "Cell-free massive MIMO with UAV access points: UAV location optimization," in *Proc. IEEE/CIC Int. Conf. Commun. China*, Aug. 2022, pp. 262–267.

[13] C. Diaz-Vilor, A. Lozano, and H. Jafarkhani, "Cell-free UAV networks: Asymptotic analysis and deployment optimization," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3055–3070, May 2023.
[14] C. Diaz-Vilor, A. Lozano, and H. Jafarkhani, "Cell-free UAV networks with wireless fronthaul: Analysis and optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2054–2069, Mar. 2024.
[15] A. Ahmed, C. Chaieb, W. Ajib, H. Elbiaze, and R. Glitho, "Resource allocation and UAVs placement in cell-free wireless networks," in *Proc. IEEE Glob. Commun. Conf.*, Dec. 2022, pp. 5995–6000.
[16] C. Liu, W. Feng, Y. Chen, C.-X. Wang, and N. Ge, "Cell-free satellite-UAV networks for 6G wide-area Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1116–1131, Apr. 2021.
[17] Z. Wu and Q. Wang, "Trajectory optimization and power allocation for cell-free satellite-UAV Internet of Things," *IEEE Access*, vol. 11, pp. 203–213, 2023.
[18] Y. K. Tun, Y. M. Park, T. H. T. Le, Z. Han, and C. S. Hong, "A business model for resource sharing in cell-free UAVs-assisted wireless networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8839–8852, Aug. 2022.
[19] M. Samir, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Leveraging UAVs for coverage in cell-free vehicular networks: A deep reinforcement learning approach," *IEEE Trans. Mobile Comput.*, vol. 20, no. 9, pp. 2835–2847, Sep. 2021.
[20] D. D. Souza et al., "Effective channel DL pilot-based estimation in user-centric cell-free massive MIMO networks," in *Proc. IEEE Glob. Commun. Conf.*, Dec. 2022, pp. 705–710.
[21] G. Femenias and F. Riera-Palou, "Fronthaul-constrained cell-free massive MIMO with low resolution ADCs," *IEEE Access*, vol. 8, pp. 116195–116215, 2020.
[22] O. T. Demir, M. Masoudi, E. Björnson, and C. Cavdar, "Cell-free massive MIMO in O-RAN: Energy-aware joint orchestration of cloud, fronthaul, and radio resources," *IEEE J. Sel. Areas Commun*, vol. 42, no. 2, pp. 356–372, Feb. 2024.
[23] A. L. P. Fernandes et al., "A cost assessment methodology for user-centric distributed massive MIMO architectures," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 3517–3543, 2024.
[24] Y. Chen, X. Zhang, F. Yao, K. An, G. Zheng, and S. Chatzinotas, "Pilot assignment and power control in secure UAV-enabled cell-free massive MIMO networks," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 3377–3391, Jan. 2024.
[25] C. D'Andrea, A. Garcia-Rodriguez, G. Geraci, L. G. Giordano, and S. Buzzi, "Cell-free massive MIMO for UAV communications," in *Proc. IEEE Int. Conf. Commun. Workshops*, May 2019, pp. 1–6.
[26] S.-N. Jin, D.-W. Yue, and H. H. Nguyen, "Spectral and energy efficiency in cell-free massive MIMO systems over correlated Rician fading," *IEEE Syst. J.*, vol. 15, no. 2, pp. 2822–2833, Jun. 2021.
[27] A. A. Polegre, F. Riera-Palou, G. Femenias, and A. G. Armada, "Channel hardening in cell-free and user-centric massive MIMO networks with spatially correlated Ricean fading," *IEEE Access*, vol. 8, pp. 139827–139845, 2020.
[28] 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.
[29] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency (Foundations and Trends in Signal Processing Series)*. Delft, The Netherlands: Now Publishers, 2018.
[30] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
[31] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inf. Theory.*, vol. 6, no. 1, pp. 7–12, Mar. 1960.
[32] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, vol. 159. Berlin, Germany: Springer Science & Business Media, 2012.
[33] M. M. M. Freitas et al., "Scalable user-centric distributed massive MIMO systems with limited processing capacity," in *Proc. IEEE Int. Conf. Commun.*, 2023, pp. 4298–4304.
[34] S. Malkowsky et al., "The world's first real-time testbed for massive MIMO: Design, implementation, and validation," *IEEE Access*, vol. 5, pp. 9073–9088, 2017.
[35] C. Desset and B. Debaillie, "Massive MIMO for energy-efficient communications," in *Proc. 46th Eur. Microw. Conf.*, Oct. 2016, pp. 138–141.
[36] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *Proc. IEEE 81st Veh. Technol. Conf.*, May 2015, pp. 1–7.
[37] 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, Jun. 2015.
[38] 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, Mar. 2018.
[39] K. Dorling, J. Heinrichs, G. G. Messier, and S. Magierowski, "Vehicle routing problems for drone delivery," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 1, pp. 70–85, Jan. 2017.
[40] J. Zhang, J. F. Campbell, D. C. Sweeney II, and A. C. Hupman, "Energy consumption models for delivery drones: A comparison and assessment," *Transp. Res. Part D, Transp. Environ.*, vol. 90, Jan. 2021, Art. no. 102668.
[41] 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, Nov. 2019.

Publication V

Freitas, M. M. M., Souza, D. D., Fernandes, A. L. P., Costa, D. B. da,
Mendes Cavalcante, A., Valcarenghi, L., and Weyl Albuquerque Costa, J. C.
Scalable User-Centric Distributed Massive MIMO Systems With Restricted Processing Capacity

Reprinted with permission from
IEEE Transactions on Wireless Communications
Vol. 23, no. 12, pp. 19933 – 19949, December 2024

Scalable User-Centric Distributed Massive MIMO Systems With Restricted Processing Capacity

Marx M. M. Freitas, Daynara D. Souza, André L. P. Fernandes, Daniel Benevides da Costa, *Senior Member, IEEE*, André Mendes Cavalcante, *Member, IEEE*, Luca Valcarenghi, *Senior Member, IEEE*, and João C. Weyl Albuquerque Costa, *Senior Member, IEEE*

***Abstract*— This paper investigates the performance of scalable user-centric (UC) distributed massive multiple-input multiple-output (D-mMIMO) systems with multiple central processing units (CPUs), commonly called cell-free mMIMO. Specifically, a framework incorporating processing capacity and inter-CPU communication constraints is proposed. Two methods are presented for limiting the number of radio units (RUs) serving each user equipment (UE). The first method is performed by the CPUs, while the second one is implemented at the UEs and RUs. Both methods prevent the computational complexity (CC) for channel estimation and precoding signals from increasing with the number of RUs. The backhaul signaling demands are presented and modeled, and it is considered that each CPU can serve only a restricted number of UEs managed by other CPUs to mitigate inter-CPU communication. Two strategies to adjust the RU clusters according to the network implementations are also proposed. We compare the proposed approaches with a traditional scalable UC system. Simulation results reveal that the proposed techniques allow UC systems to keep their spectral efficiency (SE) under minor degradation while reducing the CC by 98% and improving energy efficiency (EE). Besides, managing inter-CPU communication controls backhaul traffic effectively, and RU cluster adjustments further reduce CC.**

***Index Terms*— Cell-free networks, computational complexity, multiple CPUs, RU selection, user-centric approach.**

Received 11 January 2024; revised 13 August 2024; accepted 25 October 2024. Date of publication 12 November 2024; date of current version 12 December 2024. This work was supported in part by the Innovation Center, Ericsson Telecomunicações Ltda., Brazil; in part by the National Council for Scientific and Technological Development (CNPq); and in part by the project Collaborative edge-cLoud continuum and Embedded AI for a Visionary industry of thE futuRe (CLEVER) funded by the Key Digital Technologies Joint Undertaking and its members [including top-up funding by the Italian Ministry of Research and University (MUR)] under Project 101097560. An earlier version of this paper was presented in part at the IEEE International Conference on Communications (ICC), in 2023 [DOI: 10.1109/ICC45041.2023.10279694]. The associate editor coordinating the review of this article and approving it for publication was G. Wei. *(Corresponding author: Daniel Benevides da Costa.)*

Marx M. M. Freitas and André L. P. Fernandes are with the Applied Electromagnetism Laboratory, Federal University of Pará (UFPA), Belém 66075-110, Brazil (e-mail: marx@ufpa.br; andrelpf@ufpa.br).

Daynara D. Souza is with the Applied Electromagnetism Laboratory, Federal University of Pará (UFPA), Belém 66075-110, Brazil, and also with the School of Energy Systems, Lappeenranta-Lahti University of Technology (LUT), 53850 Lappeenranta, Finland (e-mail: daynara.dias.souza@lut.fi).

Daniel Benevides da Costa is with the Interdisciplinary Research Center for Communication Systems and Sensing (IRC-CSS), Department of Electrical Engineering, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia (e-mail: danielbcosta@ieee.org).

André Mendes Cavalcante is with Ericsson Research, Ericsson Telecomunicações Ltda., Indaiatuba 13337-300, Brazil (e-mail: andre.mendes.cavalcante@ericsson.com).

Luca Valcarenghi is with the Telecommunications, Computer Engineering, and Photonics Institute (TeCIP), Scuola Superiore Sant'Anna, 56127 Pisa, Italy (e-mail: luca.valcarenghi@santannapisa.it).

Joao C. Weyl Albuquerque Costa is with the Applied Electromagnetism Laboratory, Federal University of Pará (UFPA), Belém 66075-110, Brazil, and also with the Instituto Sustentabilidade da Amazônia com Ciência e Inovação (iSACI), Belém 66620-270, Brazil (e-mail: jweyl@ufpa.br).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2024.3491153.

Digital Object Identifier 10.1109/TWC.2024.3491153

I. Introduction

USER-CENTRIC (UC) distributed massive multiple-input multiple-output (D-mMIMO) systems, also referred to as cell-free (CF) mMIMO, have been envisaged as one of the most promising technologies for future mobile communication networks (6G and beyond) [1], [2], [3]. In these systems, several radio units (RUs) are spread out in the coverage area, and the user equipment (UE) is served by a subset of RUs, called RU cluster, providing a more uniform service and a better coverage probability than cell-based systems due to the enhanced macro-diversity and reduction of RU-UE distances [4], [5], [6], [7], [8], [9]. Despite the benefits, computational complexity (CC) can still be a drawback in these systems.

Several baseline solutions consider that the complexity of UC systems grows with the number of UEs and RUs, which is not practical [4], [5]. In this regard, [8], [9], and [10] proposed a framework to provide scalability to UC systems. Essentially, it limits the number of UEs each RU 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 spectral efficiency (SE) losses compared to the case when the UEs are served by all RUs. The conclusions hold for both centralized and distributed network implementations. In the former, channel estimation and combining processing tasks are carried out on the central processing units (CPUs), while in the latter, they occur on the RUs. However, although the network resources become independent of the number of UEs, the signal processing complexity can still grow with the number of RUs [9]. For instance, the number of complex multiplications required to perform channel estimation and precoding remains proportional to the number of RUs serving the UE [8]. Thus, a more in-depth investigation into this topic is necessary, as the literature regularly assumes that there are more RUs than UEs in the network.

Another limitation inherent to UC systems is that the RU selection processes are not adapted to the network implementations. They generally only intend to improve some key points, such as effective channel gain [11], reduce pilot contamination [9], among others [12], [13], [14]. Consequently, RU clusters may benefit one implementation over another. For instance, RU clusters with a large number of RUs can degrade the energy efficiency (EE) and CC of UC systems operating in distributed implementation while they can improve the SE for the centralized ones.

Most of the strategies in the literature to enhance the network performance (e.g., SE and EE) also consider that a single CPU is responsible for coordinating the signals of all RUs. In other words, those existing strategies do not evaluate the negative impacts in UC systems when employing multiple CPUs, such as increased signaling on backhaul links. That is, the CPUs may need to share signaling to serve the UE since the UE's RU cluster can comprise RUs connected to different CPUs, as illustrated in Fig. 1. This signaling demand is called inter-CPU communication or inter-CPU coordination [15], [16], [17], [18]. Furthermore, the CC required to perform channel estimation and precoding is typically modeled only for a single CPU scenario, which cannot be directly applied to multiple CPUs. Therefore, deeper investigations into these topics are indeed necessary since state-of-the-art solutions rely on UC systems composed of multiple CPUs to efficiently divide the network processing tasks [9], [10].

A. Literature Review

The UC D-mMIMO literature has proposed several approaches to reduce network complexity under computational and signaling aspects [5], [6], [7], [8], [9]. For instance, [6] introduced the UC approach, demonstrating that UC D-mMIMO systems could achieve comparable performance to canonical D-mMIMO systems while reducing CC and fronthaul requirements. In [8] and [9], the authors analyzed the scalability of D-mMIMO systems, presenting their performance in terms of SE for 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 they did not provide any analysis regarding the number of RUs. Moreover, [8] and [9] claimed that their proposed strategies are effective for UC systems with multiple CPUs but did not detail the network's necessary signaling procedures and requirements to make it successful. That is, the authors in [8] and [9] did not quantify the level of inter-CPU communication required by the network.

Strategies for reducing the number of RUs serving the UE were proposed in [12] and [13]. Nevertheless, a mechanism to prevent network processing demands from growing with the number of RUs was not presented, i.e., the maximum number of RUs serving each UE was not restricted. In [19], the maximum number of RUs 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's processing capacity limitation. In addition, a detailed investigation regarding CC and multiple CPUs was not provided.

In [15], 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 RUs inside that cell. Conversely, UEs at the cell edges can connect to RUs 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.

In [16], the UE was initially connected to a primary CPU and subsequently linked to other CPUs, referred to as non-primary CPUs. The latter designates the UE as an inter-coordinated UE. This approach effectively controlled inter-CPU signaling while keeping the SE under minor degradation. To this end, it was considered that the number of inter-coordinated UEs that each CPU serves must be restricted. However, [16] did not quantify the backhaul signaling, and the evaluations focused only on the implementations of distributed processing. Regarding the adjustment of RU clusters under different network implementations, to the best of the authors' knowledge, no other works addressing this topic were found.

B. Contributions

This paper investigates the performance of scalable UC D-mMIMO systems by assuming that the CC to perform channel estimation and precoding signals does not grow with the number of RUs. In particular, it is considered a UC system where the UE is associated only with a finite number of RUs, i.e., the UE is connected only with the RUs having the strongest channel gains. To the best of the authors' knowledge, this is the first paper to propose an approach that limits the CC of UC systems from growing with the number of RUs. Moreover, a method is proposed to adjust the RU clusters according to the network implementation. The proposed method works in UC systems with and without processing capacity limitations, and it can be used as an alternative solution for reducing CC in UC systems without processing capacity limitations. As far as the authors are aware, this is also the first work that proposes a method for adjusting the RU clusters according to the network implementation in UC systems. Moreover, the work studies the feasibility of UC systems when the signaling requirements on backhaul links do not grow with the number of UEs, i.e., the inter-communication among CPUs is controlled. Overall, the main contributions of this paper can be summarized as:

- Two strategies for controlling the RU cluster size of UEs are proposed. The first one is conducted by the CPUs, while the second strategy is performed locally between UEs and RUs. Simulation results reveal that the proposed solutions allow the SE to be kept under minor degradation even if the CC is reduced by up to 98%. However, our results also demonstrate that the centralized implementation may require more processing capacity than distributed to avoid significant losses in the SE.
- Two methods for adjusting the RU clusters according to the network implementation are proposed. The results

demonstrate that the proposed schemes can reduce CC and potentially increase EE.
- A framework is proposed to control the RU cluster size and manage signaling demands on backhaul links in each network implementation. Moreover, a model for the backhaul traffic is provided, accounting for data sharing, channel estimates, and precoding coefficients exchanged among CPUs.
- The CC is discussed by accounting for multiple CPUs, and the EE modeling is improved by considering the processing power consumption of various CPUs and backhaul links.

C. Paper Outline and Notations

The remainder of this paper is organized as follows. Section II presents the system model, including the channel estimation procedure, a framework for signaling requirements from multiple CPUs, and the downlink (DL) SE. Section III presents the modeling of CC and EE. Sections IV and V introduce the proposed approaches to limit the network processing capacity and to perform RU cluster adjustment. Section VI plots illustrative numerical results and draws insightful discussions to reveal the effectiveness of the proposed approaches compared to prior baseline strategies. Finally, Section VII concludes the paper.

Notation: Boldface lowercase and uppercase letters denote vectors and matrices, respectively. The superscript $(\cdot)^{\mathrm{H}}$ denotes the conjugate-transpose operation, the $N \times N$ identity matrix is denoted as $\mathbf{I}_N$, and the cardinality of the set $\mathcal{A}$ is represented by $|\mathcal{A}|$. The trace, euclidean norm, and expectation operator are denoted as $\mathrm{tr}(\,.\,)$, $\|\,.\,\|$, and $\mathbb{E}\{\,.\,\}$, respectively. The notation $\mathcal{N}_{\mathbb{C}}\left(\mu, \sigma^2\right)$ stands for a complex Gaussian random variable with mean μ and variance σ^2.

II. System Model

We consider a D-mMIMO network composed of J CPUs, L RUs, and K single-antenna UEs, where $L > K$. Each RU is equipped with N antennas, resulting in a total of M antennas considering all RUs, with $M = NL$. The RUs are connected to the CPUs through fronthaul links, while the CPUs are interlinked through backhaul connections, as shown in Fig. 1. The fronthaul links undergo limited transmission capacity, while the backhaul ones are considered error-free and capable of supporting the data traffic. We utilize analog-to-digital converters (ADCs) to limit the data transmitted over the fronthaul links. Therefore, signals are quantized before being sent to the fronthaul. The system operates on time-division duplex (TDD) mode and it is assumed that the uplink (UL) and DL channels are reciprocal. Thus, channel estimation is performed only in the UL direction. We focus on DL transmissions and consider that the channel $\mathbf{h}_{kl} \in \mathbb{C}^{N\times 1}$ between the RU l and UE k undergoes an independent correlated Rician fading, being defined as [20], [21], [22]

$$\mathbf{h}_{kl} = \underbrace{\sqrt{\frac{\kappa_{kl}}{1+\kappa_{kl}}}\mathbf{h}_{kl}^{\mathrm{LOS}}\, e^{j\theta_{kl}}}_{\overline{\mathbf{h}}_{kl}} + \underbrace{\sqrt{\frac{1}{1+\kappa_{kl}}}\mathbf{h}_{kl}^{\mathrm{NLOS}}}_{\widetilde{\mathbf{h}}_{kl}}, \quad (1)$$

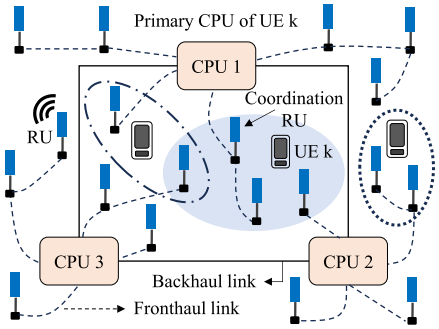


Fig. 1. UC D-mMIMO system with multiple CPUs. Each CPU is connected to a subset of RUs, and the UEs can be associated with RUs linked to different CPUs.

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 covariance matrix[1] $\widetilde{\mathbf{R}}_{kl} = \mathbb{E}\{\widetilde{\mathbf{h}}_{kl}\widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} \in \mathbb{C}^{N\times N}$. The term $\theta_{kl} \sim \mathcal{U}[0, 2\pi)$ denotes random phase shifts that may occur in LOS components due to the UEs mobility, and the Rician factor κ_{kl} is the power ratio between the LOS and non-line-of-sight (NLOS) components. The latter can be computed as $\kappa_{kl} = p_{\mathrm{LOS}}/(1 - p_{\mathrm{LOS}})$, with p_{LoS} being the probability of the LOS component's existence [24]. Furthermore, $\mathbf{h}_{kl}^{\mathrm{NLOS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{kl}^{\mathrm{NLOS}})$ represents the effects of the NLOS propagation.

Assuming that the RUs are equipped with half-wavelength-spaced uniform linear arrays (ULAs), the covariance 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 covariance presented in [23, Sec. 2.6]. Thus, the covariance matrix of the term $\widetilde{\mathbf{h}}_{kl}$ in (1) is given by $\widetilde{\mathbf{R}}_{kl} = \mathbb{E}\{\widetilde{\mathbf{h}}_{kl}\widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} = \mathbf{R}_{kl}^{\mathrm{NLOS}}/(\kappa_{kl}+1)$, 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 RU l can be expressed as [23]

$$\mathbf{h}_{kl}^{\mathrm{LoS}} = \sqrt{\beta_{kl}}\left[1, \cdots, e^{\mathrm{j}(N-1)\pi \sin(\varphi_{kl})\cos(\psi_{kl})}\right]^{\mathrm{T}}, \quad (2)$$

where φ_{kl} denotes the azimuth angle, ψ_{kl} is the elevation angle of the LOS component, and β_{kl} is the large-scale fading gain, which can be calculated as $\beta_{kl} = \mathrm{tr}(\mathbf{R}_{kl})/N$.

A. Network Implementations

UC D-mMIMO systems are commonly implemented in centralized or distributed manners according to processing capabilities. The centralized implementation places most baseband functions on the CPUs. Therefore, the CPUs are responsible for channel estimation and precoding [9]. Furthermore, they encode and quantize the DL signals. In the distributed implementation, essential processing functions, such as channel estimation, are moved to the RUs. Consequently, the CPUs are only responsible for encoding and quantization.

The centralized implementation usually offers superior interference mitigation since the CPUs can access global

[1]The statistical covariance matrix represents the large-scale fading of the system, being a function of the spatial channel covariance, path loss, antenna gains, and shadowing [23].

channel state information (CSI), which includes channel estimates and statistics. Conversely, the distributed one can be less complex and avoids the need to transmit the pilot signals on fronthaul links [7]. In this regard, several network procedures, such as channel estimation, interference mitigation, and CC, vary depending on the network implementation. In order to compute the combining and precoding vectors, this paper utilizes the partial MMSE (P-MMSE) and partial regularized zero-forcing (P-RZF) schemes for centralized implementation. For the distributed one, the local partial MMSE (LP-MMSE) and maximum ratio (MR) are utilized. These techniques have been chosen due to their scalability features [9].

B. Uplink Training and Channel Estimation

Each coherence block comprises τ_c samples, where τ_p samples are dedicated for UL pilot signals and τ_d for DL data. During the UL training phase, the UEs send pilot sequences of τ_p-length to the RUs for channel estimation. Then, the UL channels are estimated by correlating the received signals with a known pilot sequence and using phase-unaware linear minimum mean square error (LMMSE) estimation. The pilot signals are assumed to be mutually orthogonal and independent of the number of UEs K to ensure the scalability of the pilot resources. Thus, a pilot t can be reused by some UEs if the number of UEs is greater than the number of pilot signals, i.e., $K > \tau_p$. Let $\mathcal{P}_k \subset \{1, \ldots, K\}$ denote the subset of the UEs assigned to the pilot t, including the UE k. The received pilot signal at RU l can be expressed as [8]

$$\mathbf{y}_{tl}^{\text{pilot}} = \sum_{i \in \mathcal{P}_k} \sqrt{\tau_p \eta_i}\, \mathbf{h}_{il} + \mathbf{n}_{tl}, \tag{3}$$

where $\mathbf{n}_{tl} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_N, \sigma_{ul}^2 \mathbf{I}_N\right)$ denotes the noise and η_i is the power that the UE i transmits in the UL direction. The channel estimation procedure differs in each network implementation. In the distributed implementation, the channel vector $\mathbf{h}_{kl}$ is estimated locally by RU l after receiving the pilot signals $\mathbf{y}_{tl}^{\text{pilot}}$ sent by the UEs in (3). In the centralized one, the RUs forward the received pilot signals $\mathbf{y}_{tl}^{\text{pilot}}$ to the CPUs, which then perform the channel estimation. Note that due to the use of ADCs, the signals must be quantized before being sent to the fronthaul links. Therefore, in the centralized implementation, the received pilot signals $\mathbf{y}_{tl}^{\text{pilot}}$ are quantized before being sent to CPUs. In contrast, the quantization of pilot signals is not necessary for the distributed implementation, as channel estimation occurs locally in the RUs.

In the distributed implementation, the LMMSE channel estimation is given by [8]

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p \eta_k} \mathbf{R}_{kl} \boldsymbol{\Psi}_{tl}^{-1} \mathbf{y}_{tl}^{\text{pilot}}, \tag{4}$$

where $\mathbf{R}_{kl} = \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\text{H}}\} = (\overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\text{H}} + \widetilde{\mathbf{R}}_{kl})$ and $\boldsymbol{\Psi}_{tl} = \mathbb{E}\{(\mathbf{y}_{tl}^{\text{pilot}})(\mathbf{y}_{tl}^{\text{pilot}})^{\text{H}}\} = \sum_{i \in \mathcal{P}_k} \eta_i \tau_p (\overline{\mathbf{h}}_{il}\overline{\mathbf{h}}_{il}^{\text{H}} + \widetilde{\mathbf{R}}_{il}) + \sigma_{\text{ul}}^2 \mathbf{I}_N$. The term $\boldsymbol{\Psi}_{tl}$ denotes the covariance matrix of the received signal $\mathbf{y}_{tl}^{\text{pilot}}$. One can note that $\boldsymbol{\Psi}_{tl}$ also indicates the presence of pilot contamination since it contains the sum of the covariance matrices of all UEs sharing pilot t.

In the centralized implementation, the LMMSE channel estimation is computed as [25]

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p \eta_k} \mathbf{R}_{kl} \widetilde{\boldsymbol{\Psi}}_{tl}^{-1} \widetilde{\mathbf{y}}_{tl}^{\text{pilot}}, \tag{5}$$

where $\widetilde{\mathbf{y}}_{tl}^{\text{pilot}} \approx \alpha_{\text{p},l} \mathbf{y}_{tl}^{\text{pilot}} + \boldsymbol{q}_t$ stands for the quantized pilot signal received on the CPU connected to RU l, while $\widetilde{\boldsymbol{\Psi}}_{tl} \approx \alpha_{\text{p},l}^2 \boldsymbol{\Psi}_{tl} + \alpha_{\text{p},l}\left(1 - \alpha_{\text{p},l}\right) \boldsymbol{\Psi}_{tl}$ denotes the covariance matrix of $\widetilde{\mathbf{y}}_{tl}^{\text{pilot}}$. The term $\alpha_{\text{p},l}$ represents a distortion factor associated with the number of bits b_l^{p} used to quantize the pilot signals, and $\boldsymbol{q}_t \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{\boldsymbol{q}_t}) \in \mathbb{C}^{N \times 1}$ means the quantization noise with covariance $\boldsymbol{R}_{\boldsymbol{q}_t} = \alpha_{\text{p},l}\left(1 - \alpha_{\text{p},l}\right) \boldsymbol{\Psi}_{tl}$. Note that (5) degenerates to (4) when $\alpha_{\text{p},l} = 1$. The relationship between $\alpha_{\text{p},l}$ and b_l^{p} is obtained from [26] and [27].

C. Downlink Data Transmission

In UC systems, each UE is associated with a subset of RUs called RU cluster. These clusters are intended to be dynamic and capable of adapting to variations in network conditions, such as UE position and channel properties. To represent the RU clusters of each UE, we proceed as follows: First, let $\mathcal{M}_k \subset \{1, \ldots, L\}$ represent the indexes of the RUs serving the UE k. Second, let vector $\mathbf{c}_k = [\text{c}_{k1}, \ldots, \text{c}_{kL}] \in \mathbb{N}^{1 \times L}$ denote the RUs that establish a connection with UE k, such that $\text{c}_{kl} = 1$ if the RU serves the UE k, and $\text{c}_{kl} = 0$ otherwise. Therefore, the connections between the UE k and RUs are expressed as

$$\text{c}_{kl} = \begin{cases} 1 & \text{if } l \in \mathcal{M}_k \\ 0 & \text{if } l \notin \mathcal{M}_k. \end{cases} \tag{6}$$

The matrix $\mathbf{D}_{kl} \in \mathbb{N}^{N \times N}$ is also utilized to describe which antennas of the RU l serve the UE k. It is assumed that all N antennas of RU l serve the UE k; thus $\mathbf{D}_{kl} = \mathbf{I}_N$ when $\text{c}_{kl} = 1$. Otherwise, $\mathbf{D}_{kl} = \mathbf{0}_N$. The vector $\mathbf{c}_{kl}$ can also be utilized to compute the number of UEs that RU l serves and the number of RUs serving the UE k. For instance, let $\mathcal{D}_l$ represent the indexes of the subset of UEs that RU l serves. The cardinalities of $\mathcal{D}_l$ and $\mathcal{M}_k$ can be computed as $|\mathcal{D}_l| = \sum_{k \in \mathcal{D}_l} \text{c}_{kl}$ and $|\mathcal{M}_k| = \sum_{l \in \mathcal{M}_k} \text{c}_{kl}$. These cardinalities can also be represented by L_k and K_l, where $L_k = |\mathcal{M}_k|$ and $K_l = |\mathcal{D}_l|$. It is noteworthy that scalable UC D-mMIMO systems usually assume that K_l is constrained to prevent it from being a function of the number of UEs. Thus, it is assumed that $K_l \leq \tau_p$ [9].

Once the RU clusters are formed, the network can proceed with the processes of channel estimation, precoding, and DL transmission for the UEs that each RU serves. Let $\mathbf{x}_l^{\text{dist}} = \sum_{i=1}^{K} \alpha_{il} \mathbf{D}_{il} \mathbf{w}_{il}\left(s_i + q_{il}\right)$ and $\mathbf{x}_l^{\text{cent}} = \alpha_l \sum_{i=1}^{K} \mathbf{D}_{il} \mathbf{w}_{il} s_i + \boldsymbol{q}_l$ denote the data signals sent by RU l in the distributed and centralized implementations. Here, $s_i \in \mathbb{C}$ is the unity-power symbol intended for UE i, and α_{il} and α_l represent the distortion factors associated with the number of bits used to quantize the DL signals in distributed and centralized implementations, respectively. Besides, $q_{il} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{q_{il}}^2)$ and $\boldsymbol{q}_l \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{\boldsymbol{q}_l}) \in \mathbb{C}^{N \times 1}$ are the quantization noises applied to the DL signals, with covariances $\sigma_{q_{il}}^2 =$

$\alpha_{il}\left(1-\alpha_{il}\right)$ and $\boldsymbol{R}_{q_l} = \alpha_l\left(1-\alpha_l\right)\sum_{i\in\mathcal{M}_k}\mathbb{E}\left\{\mathbf{w}_{il}\mathbf{w}_{il}^{\mathrm{H}}\right\}$. An achievable DL SE can be computed as [9], [25]

$$\mathrm{SE}_k^{(\mathrm{dl})} = \frac{\tau_d}{\tau_c}\log_2\left(1+\frac{\mathrm{DS}_k}{\mathrm{IS}_k-\mathrm{DS}_k+\mathrm{QN}_k+\sigma_{dl}^2}\right), \quad (7)$$

where $\mathrm{IS}_k = \sum_{i=1}^{K}\mathbb{E}\{\left|\sum_{l=1}^{L}\widetilde{\alpha}_{il}\mathbf{h}_{kl}^{\mathrm{H}}\mathbf{D}_{il}\mathbf{w}_{il}\right|^2\}$ stands for the interference, $\mathrm{DS}_k = \left|\sum_{l=1}^{L}\mathbb{E}\{\widetilde{\alpha}_{kl}\mathbf{h}_{kl}^{\mathrm{H}}\mathbf{D}_{kl}\mathbf{w}_{kl}\}\right|^2$ denotes the desired signal, QN_k means the quantization noise, and σ_{dl}^2 is the receiver noise variance. One can note that $\widetilde{\alpha}_{il}$ is the distortion factor associated with the number of bits used to quantize the DL signals. Both $\widetilde{\alpha}_{il}$ and QN_k are computed differently in each network implementation. Thus, $\widetilde{\alpha}_{il} = \alpha_{il}$ and $\widetilde{\alpha}_{il} = \alpha_l$ for distributed and centralized implementations, respectively, whereas QN_k is given by

$$\mathrm{QN}_k = \begin{cases} \mathbb{E}\left\{\left|\sum_{l=1}^{L}\mathbf{h}_{kl}^{\mathrm{H}}\mathbf{D}_{kl}\boldsymbol{q}_l\right|^2\right\}, & \text{for CI} \\ \mathbb{E}\left\{\left|\sum_{l=1}^{L}\mathbf{h}_{kl}^{\mathrm{H}}\sum_{i=1}^{K}\mathbf{D}_{il}\mathbf{w}_{il}q_{il}\right|^2\right\}, & \text{for DI.} \end{cases} \quad (8)$$

where CI and DI represent the centralized and distributed implementations, respectively. The term $\mathbf{w}_{il}$ is the precoding vector generated to mitigate the inferences between the UE i and RU l. In the distributed implementation, $\mathbf{w}_{il}$ satisfies $\mathbb{E}\{\left\|\mathbf{w}_{il}\right\|^2\} = \rho_{il}$, with ρ_{il} being the power allocated to the UE i regarding the RU l. In the centralized one, the fraction of power assigned to $\mathbf{w}_{il}$ is calculated based on the collective precoding vector $\mathbf{w}_i = \left[\mathbf{w}_{i1}^{\mathrm{T}},\ldots,\mathbf{w}_{iL}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$, which must satisfy $\mathbb{E}\{\left\|\mathbf{w}_i\right\|^2\} = \rho_i$, where ρ_i represents the transmit power assigned to UE i by all its serving RUs.

D. CPUs Requirements and Backhaul Signaling

In UC D-mMIMO systems involving multiple CPUs, it is essential to note that the RU cluster serving a specific UE may comprise RUs connected to different CPUs, as depicted in Fig. 1. Consequently, signal processing tasks must be efficiently distributed among these CPUs [7], [9], [10], [18]. This paper considers two classes of CPUs within the network to distribute the processing load among them more effectively. Specifically, one of the CPUs associated with the UE's RU cluster is designated as its primary CPU, i.e., the one responsible for the majority of its signal processing tasks. The remaining CPUs serving the UE are called secondary CPUs [16]. It is assumed that each UE is associated with a primary CPU, which is one of the CPUs belonging to the UE's RU cluster. Therefore, UEs may be assigned to different primary CPUs, since the UEs can be positioned in distinct positions in the coverage area.

This section fills a gap in the existing literature by providing a framework explaining how the CPUs should operate in each network implementation and highlighting the signaling requirements each implementation can impose on backhaul links. Moreover, a detailed modeling for the signaling demands on backhaul links, which considers the impacts of instantaneous CSI and data sharing, is also provided.

1) Centralized Implementation: Each CPU belonging to the RU cluster of UE k estimates the channel vector $\widehat{\mathbf{h}}_{kl}$ for all its RUs serving the UE k. As for the primary CPU of UE k, it also computes the combining and precoding vectors of UE k. Hereinafter, the primary CPU of UE k will also be denoted as CPU j_k. It is considered that the primary CPU of UE k is the CPU connected to the RU with the strongest channel gain serving the UE k [16]. The procedure for associating a primary CPU with a UE is detailed in Subsection IV-A.

Regarding the signaling demands, the primary CPU of UE k requests other CPUs in the network to forward the channel estimates of the UE k and interfering UEs to compute the combining and precoding ($\mathbf{w}_k$) vectors. After generating $\mathbf{w}_k$, the CPU j_k sends $\mathbf{w}_k$ and s_k to the secondary CPUs. Then, the secondary CPUs of UE k quantize and forward the DL data s_k to their respective RUs. Let $\mathcal{J}_k^{\mathrm{sec}}$ denote the subset of secondary CPUs associated with the RU cluster of UE k. The number of complex scalars that the secondary CPUs have to exchange with the primary CPU of UE k via backhaul in each coherence block can be modeled as

$$\mathrm{BT}_k^{\mathrm{pri}} = \sum_{j\in\mathcal{J}_k^{\mathrm{sec}}}\left(2L_{jk}N+\tau_d\right), \quad (9)$$

where L_{jk} denotes the number of RUs in the secondary CPU j serving the UE k. Moreover, L_{jk} is multiplied by 2 in (9) to account for the transmission of channel estimates and precoding vectors over backhaul links. Similarly, the number of complex scalars exchanged via backhaul between a subset of CPUs and the primary CPU of UE k to transmit channel estimates of interfering UEs is given by

$$\mathrm{BT}_k^{\mathrm{int}} = \sum_{i\neq k, i\in\mathcal{I}_k}\sum_{j'\in\mathcal{J}_i}L_{j'i}N, \quad (10)$$

where $\mathcal{I}_k$ denotes the subset of interfering UEs affecting the signal of UE k and $\mathcal{J}_i$ represents the subset of CPUs associated with each interfering UE i, with $j'\neq j_k$. One can note that (10) may overestimate the backhaul traffic, as channel estimates from interfering UEs are sent redundantly to CPU j_k, i.e., they are transmitted for each UE utilizing CPU j_k as its primary CPU. However, there is no need for redundant transmission of channel estimates for these interfering UEs to CPU j_k since the UEs that use CPU j_k as their primary CPU may share the same interfering UEs. In this case, these channel estimates can be transmitted only once. Thus, (10) can be rewritten as

$$\mathrm{BT}_k^{\mathrm{int}} = \sum_{i\neq k, i\in\widetilde{\mathcal{I}}_k}\sum_{j'\in\mathcal{J}_i}L_{j'i}N, \quad (11)$$

where $j'\neq j_k$ and $\widetilde{\mathcal{I}}_k$ represents the subset of interfering UEs whose channel estimates have not yet been sent to CPU j_k. Hence, the total number of complex scalars exchanged on the backhaul links is calculated as

$$\mathrm{BT}_k^{\mathrm{cent}} = \mathrm{BT}_k^{\mathrm{pri}} + \mathrm{BT}_k^{\mathrm{int}}. \quad (12)$$

2) Distributed Implementation: The primary CPU of UE k encodes the DL data. Then, it sends the DL data to the secondary CPUs. The latter quantizes and forwards s_k to their respective RUs for DL data precoding. Thus, the total number

of complex scalars exchanged between the primary CPU of UE k and the secondary CPUs is given by

$$\mathrm{BT}_k^{\mathrm{dist}} = \sum_{j \in \mathcal{J}_k^{\mathrm{sec}}} \tau_d. \tag{13}$$

Remark 1: The computation of (9) and (11) relies on the assumption that the CPUs share the RU clusters of the UEs with each other. For instance, CPU j reports which RUs connected to it serve the UE k. Thus, CPU j does not need to send $L \times N$ complex scalars (to share the estimated channel and precoding of UE k) to other CPUs. Instead, it only transmits $L_{jk} \times N$ elements. This assumption is reasonable as long as the formation of RU clusters is based on channel statistics. Hence, the signaling required to share the RU clusters across the backhaul links is negligible, as the channel statistics are constant throughout the data transmission.[2]

E. Reducing Inter-CPU Communication

To further reduce the number of complex scalars exchanged between CPUs, we rely on one of our previous works [16], which proposes a strategy to mitigate the inter-communication between CPUs, also called inter-CPU coordination. Specifically, it is considered that the network has two classes of UEs. The first one comprises the UEs using the CPU j as a primary CPU, denoted as $\mathcal{K}_j^{\mathrm{pri}}$ in this paper. The second one includes the UEs that utilize CPU j as a secondary CPU, represented by subset $\mathcal{K}_j^{\mathrm{sec}}$. The approach proposed in [16] states that each CPU can serve only a limited number of UEs as a secondary CPU, denoted as K^{sec}. Thus, the CPUs associated with UE k can modify its RU cluster to meet this condition, i.e., $\left|\mathcal{K}_j^{\mathrm{sec}}\right| \leq K^{\mathrm{sec}}$. Let $\mathbf{e}_k = [\mathrm{e}_{k1}, \ldots, \mathrm{e}_{kL}] \in \mathbb{N}^{1\times L}$ represent the RU cluster of UE k. The CPUs modify $\mathbf{e}_k$ to

$$\mathbf{c}_k = \mathbf{e}_k \wedge \mathbf{f}_k, \tag{14}$$

where $\wedge$ is the logical operation AND. Besides, $\mathbf{f}_k = [\mathrm{f}_{k1}, \ldots, \mathrm{f}_{kL}] \in \mathbb{N}^{1\times L}$ is a fine-tuned version of the RU cluster of UE k, which can be expressed as

$$\mathrm{f}_{kl} = \begin{cases} 1 & \text{if } \left(k \in \mathcal{K}_j^{\mathrm{pri}}\right) \vee \left(\left|\mathcal{K}_j^{\mathrm{sec}}\right| < K^{\mathrm{sec}}\right) \\ 1 & \text{if } \left(\left|\mathcal{K}_j^{\mathrm{sec}}\right| = K^{\mathrm{sec}}\right) \wedge \left(\mathrm{G}_{kj} > \mathrm{G}_{(i_{\min})j}\right) \\ 0 & \text{otherwise} \end{cases}, \tag{15}$$

where $\vee$ is the logical operation OR. The term $\mathrm{G}_{kj} = \sum_{l \in \mathcal{L}_{kj}} \beta_{kl}$ represents the partial sum gain, with $\mathcal{L}_{kj}$ being the subset of RUs serving the UE k that are connected to CPU j, and $i_{\min}$ standing for the UE using the CPU j as a secondary CPU presenting the smallest partial sum gain. It is noteworthy that the UE $i_{\min}$ is dropped from all RUs connected to CPU j if $\mathrm{G}_{kj} > \mathrm{G}_{(i_{\min})j}$. Note that (14) limits the number of UEs using CPU j as a secondary CPU, since $\left|\mathcal{K}_j^{\mathrm{sec}}\right| \leq K^{\mathrm{sec}}$. Therefore, this paper models the total number of complex scalars that the secondary CPUs have to exchange with primary CPUs through the backhaul as

$$\mathrm{BT}^{\mathrm{pri}} = \sum_{j \in \mathcal{J}^{\mathrm{sec}}} \sum_{k \in \mathcal{K}_j^{\mathrm{sec}}} \tau_d + \mathbb{I}\left(2L_{jk}N\right), \tag{16}$$

where the binary indicator $\mathbb{I} \in \{0, 1\}$ specifies the network implementation type, with $\mathbb{I} = 1$ corresponding to the centralized implementation, and $\mathbb{I} = 0$ to the distributed one. The term $\mathcal{J}^{\mathrm{sec}}$ denotes the subset of CPUs acting as a secondary CPU. It can be seen that the condition $\left|\mathcal{K}_j^{\mathrm{sec}}\right| \leq K^{\mathrm{sec}}$ prevents (16) from growing with the number of UEs K. That is, even if the number of UEs K goes to infinity, the CPU j will serve at most K^{sec} UEs as a secondary CPU. In other words, the number of sums performed in $\sum_{k \in \mathcal{K}_j^{\mathrm{sec}}} \tau_d + \mathbb{I}\left(2L_{jk}N\right)$ will be upper-bounded by K^{sec}, since $\left|\mathcal{K}_j^{\mathrm{sec}}\right| \leq K^{\mathrm{sec}}$. This condition is not met in a traditional UC system since $\left|\mathcal{K}_j^{\mathrm{sec}}\right|$ is not upper-bounded by K^{sec}. Instead, it would be a function of the number of UEs, such that $\left|\mathcal{K}_j^{\mathrm{sec}}\right| \leq K$. For instance, in the worst-case scenario, a CPU j could serve all UEs as a secondary CPU, leading to the second summation of (16) becoming $\sum_{k=1}^{K} \tau_d + \mathbb{I}\left(2L_{jk}N\right)$.

Moreover, one can note that $\mathrm{BT}^{\mathrm{pri}}$ can be further reduced by limiting the number of RUs serving the UE k to a maximum value called C_{max}, such that $L_k \leq \mathrm{C}_{max}$. Section IV discusses how to compute C_{max} in more detail. Finally, note that (16) is equivalent to $\mathrm{BT}^{\mathrm{pri}} = \sum_{k=1}^{K} \mathrm{BT}_k^{\mathrm{dist}}$ for $\mathbb{I} = 0$, and $\mathrm{BT}^{\mathrm{pri}} = \sum_{k=1}^{K} \mathrm{BT}_k^{\mathrm{pri}}$ for $\mathbb{I} = 1$.

F. Required Fronthaul and Backhaul Bit Rates

The fronthaul bit rate required by UC D-mMIMO systems varies depending on the network implementation. In the distributed implementation, the bit rate scales with the number of UEs served by the RU (i.e., K_l). In the centralized one, it scales with the number of antennas N deployed at RU. The required fronthaul traffic (in bit/s) can be expressed as [25]

$$R_{\mathrm{fh},l} = 2B\frac{\tau_d}{\tau_c}\left(N\left[b_l + b_l^{\mathrm{p}}\frac{\tau_p}{\tau_d}\right]\mathbb{I} + \sum_{k \in \mathcal{D}_l} b_{kl}\left(1 - \mathbb{I}\right)\right), \tag{17}$$

where B represents the system bandwidth, while b_l and b_{kl} are the number of quantization bits per sample used for DL data transmission in centralized and distributed implementations, respectively. In addition, b_l^{p} is the number of quantization bits per sample used for pilot signals. One can note that the fronthaul link requirements are constrained by the number of quantization bits for $\mathbb{I} = 1$. Conversely, fronthaul requirements may rise with the number of UEs served by RU l (K_l) when $\mathbb{I} = 0$. However, the maximum value of $R_{\mathrm{fh},l}$ remains constant in a scalable system, i.e., $K_l \leq \tau_p$. The bit rate R_{bh} exchanged in all backhaul links is modeled in this paper as

$$R_{\mathrm{bh}} = \frac{2B}{\tau_c}\left(\sum_{k=1}^{K} b_k^{\mathrm{bh}}\left[\left(\mathrm{BT}_k^{\mathrm{cent}}\right)\mathbb{I} + \mathrm{BT}_k^{\mathrm{dist}}(1 - \mathbb{I})\right]\right), \tag{18}$$

where b_k^{bh} represents the number of bits utilized to quantize the signals traveling in the backhaul links. It is assumed that $b_k^{\mathrm{bh}} = b_l^{\mathrm{p}} = b^{\max}$, with $b^{\max}$ being the maximum number of quantization bits per sample. The purpose is to transmit

[2] In practice, channel statistics can change due to UE mobility or scheduling. However, a deeper investigation into this topic is out of the scope of this paper.

pilots and signals on backhaul links at maximum resolution to reduce quantization errors in channel estimation and data received on secondary CPUs. The number of quantization bits per sample utilized for DL data transmission (b_{kl}) is computed by considering that all RUs are operating at their maximum capacity, i.e., $K_l = \tau_p$. Thus, b_{kl} can be computed from (17) for $\mathbb{I} = 0$ as $b_{kl} = \left\lfloor R_{\text{fh},l}^{\max} \tau_c / 2B\tau_d\tau_p \right\rfloor$ for all k in $\mathcal{D}_l$, where $\lfloor . \rfloor$ is the floor operation, and $R_{\text{fh},l}^{\max}$ denotes the maximum transmission capacity of each fronthaul link. Similarly, b_l can be calculated as $b_l = \lfloor R_{\text{fh},l}\tau_c/(2B\tau_d N) - b_l^{\text{p}}(\tau_p/\tau_d) \rfloor$. This paper assumes that $R_{\text{fh},l}^{\max} = 10\,\text{Gbps}$ and $b^{\max} = 12$. It is also considered that $b_l = b_{kl}$ to perform a fair comparison between centralized and distributed implementations.[3] Therefore, $b_l = b_{kl} = \min(b_l, b_{kl})$.

III. Computational Complexity and Energy Efficiency

A. Computational Complexity

The CC required for signaling processing tasks differs between network implementations for both CPUs and RUs. Therefore, this section also utilizes the binary indicator $\mathbb{I} \in \{0, 1\}$ to distinguish them. The CC required from each CPU j in giga operations per second (GOPS) can be expressed as

$$CC_{\text{CPU},j} = S_f \left(CC_{\text{CPU},j}^{\text{cecb}} + CC_{\text{CPU},j}^{\text{rcp}} \right) \mathbb{I} + CC_{\text{CPU},j}^{\text{basic}}, \tag{19}$$

where $CC_{\text{CPU},j}^{\text{cecb}} = CC_{\text{CPU},j}^{\text{est}} + CC_{\text{CPU},j}^{\text{comb}}$ denotes the number of complex multiplications that CPU j needs to perform channel estimation and generate the combining vectors. $CC_{\text{CPU},j}^{\text{rcp}} = (1+\tau_d) N \sum_{l \in \mathcal{J}_l} K_l$ stands for the CC associated with reciprocity calibration and precoding, with $\mathcal{J}_l$ being the subset of RUs connected to CPU j. Additionally, $S_f = 8N_{sc}/T_s\tau_c 10^9$ is a scaling factor that converts $CC_{\text{CPU},j}^{\text{cecb}}$ and $CC_{\text{CPU},j}^{\text{rcp}}$ into GOPS, where T_s is the orthogonal frequency-division multiplexing (OFDM) symbol duration, and N_{sc} represents the number of subcarriers. The last term of (19) is the CC in GOPS associated with higher-layer control/network functions, channel coding, mapping/demapping, and OFDM modulation/demodulation. These are computed following [28].

It is worth mentioning that $CC_{\text{CPU},j}^{\text{est}}$ and $CC_{\text{CPU},j}^{\text{comb}}$ are obtained from Table I, where $\mathcal{S}_k = \{i : \mathbf{D}_k \mathbf{D}_i \neq \mathbf{0}_{LN \times LN}\}$ represents the subset of UEs that are partially served by the same RUs as UE k. The term $\mathcal{K}_j^{\text{all}}$ denotes the subset of all UEs that CPU j is serving, thus $\mathcal{K}_j^{\text{all}} = \mathcal{K}_j^{\text{pri}} \cup \mathcal{K}_j^{\text{sec}}$. The CC required from RU l in GOPS can be computed as [29]

$$CC_{\text{RU},l} = S_f \left(CC_{\text{RU},l}^{\text{cecb}} + CC_{\text{RU},l}^{\text{rcp}} \right) (1 - \mathbb{I}) + CC_{\text{RU},l}^{\text{other}}, \tag{20}$$

where $CC_{\text{RU},l}^{\text{cecb}} = CC_{\text{RU},l}^{\text{est}} + CC_{\text{RU},l}^{\text{comb}}$ denotes the number of complex multiplications required by RU l to perform channel estimation and generate the combining vectors. $CC_{\text{RU},l}^{\text{rcp}} = (1+\tau_d) N K_l$ is the CC associated with reciprocity calibration and precoding application. The specific values of $CC_{\text{RU},l}^{\text{est}}$ and $CC_{\text{RU},l}^{\text{comb}}$ are calculated in Table I. The last term of (20) is obtained as $CC_{\text{RU},l}^{\text{other}} = CC^{\text{DFT}} + CC^{\text{bbf}}$, where $CC^{\text{DFT}} = 8NN_{DFT}\log_2(N_{DFT})/T_s 10^9$ is the CC in GOPS due to discrete Fourier transform (DFT) operations, with $N_{DFT} \leq N_{sc}$ being the dimension of the DFT [30]. We have assumed that $N_{DFT} = N_{sc}$. Moreover, $CC^{\text{bbf}} = 40Nf_s/10^9$ represents the CC in GOPS related to baseband filtering, considering a filter with ten taps in a polyphase filtering scheme, where f_s is the sampling frequency [31].

B. Energy Efficiency

The energy efficiency (EE) in bit/Joule is calculated as the ratio between the sum throughput in bit/s and the total power consumed in Watts (W), being expressed as

$$\text{EE}_{\text{tot}} = \frac{B \sum_{k=1}^{K} \text{SE}_k}{\sum_{l=1}^{L} \{P_l + P_{\text{fh},l}\} + P_{\text{CPUs}}^{\text{proc}} + P_{\text{bh}}}, \tag{21}$$

where P_l denotes the total power consumption in RU l, while $P_{\text{fh},l}$ represents the power consumed by the fronthaul link connected to RU l. Additionally, $P_{\text{CPUs}}^{\text{proc}}$ is the power that all CPUs need for processing tasks, and P_{bh} accounts for power consumption in all backhaul links. P_l is calculated as $P_l = \mathbb{E}\{ \|\mathbf{x}_l\|^2 \}/\gamma_l + NP_{\text{tc},l} + P_{\text{RU},l}^{\text{proc}}$, where $0 < \gamma_l \leq 1$ represents the efficiency of the power amplifier, $P_{\text{tc},l}$ is the power required for each antenna of RU l to operate internal components like converters and filters, and $P_{\text{RU},l}^{\text{proc}}$ accounts for the power needed by RU l to perform processing tasks. The latter can be given by [32]

$$P_{\text{RU},l}^{\text{proc}} = P_{\text{RU},0}^{\text{proc}} + \Delta_{\text{RU},l}^{\text{proc}} \left(\frac{CC_{\text{RU},l}}{CC_{\text{RU}}^{\max}} \right), \tag{22}$$

where $P_{\text{RU},0}^{\text{proc}}$ is the power consumed by each digital signal processor (DSP) of RU l in idle mode; $\Delta_{\text{RU},l}^{\text{proc}}$ is the slope of power consumption due to processing in RU l, and $CC_{\text{RU}}^{\max}$ indicates the maximum GOPS capacity of the DSP in RU l.

The power consumed in each fronthaul link is calculated as $P_{\text{fh},l} = P_{0,l} + P_{\text{ft},l}, R_{\text{fh},l}$, where $P_{0,l}$ is the fixed power consumption of each fronthaul link, $P_{\text{ft},l}$ denotes the traffic-dependent power in Watt per bit/s, and $R_{\text{fh},l}$ is computed in (17). Similarly, for backhaul links, $P_{\text{bh}} = 0.5 \times J(J-1)P_{\text{bh},0} + P_{\text{bt}}R_{\text{bh}}$, where $P_{\text{bh},0}$ and P_{bt} represent the fixed and traffic-dependent power of each backhaul link, while R_{bh} is computed in (18). The term $0.5 \times J(J-1)$ refers to a fully connected topology, where each CPU has a direct connection to each other. Finally, $P_{\text{CPUs}}^{\text{proc}}$ can be expressed as

$$P_{\text{CPUs}}^{\text{proc}} = \frac{1}{\sigma_{\text{cool}}} \left(\Delta_{\text{GPP}}^{\text{proc}} \frac{CC_{\text{CPUs}}}{CC_{\text{GPP}}^{\max}} + \chi_{\text{CPUs}}^{\text{proc}} \right), \tag{23}$$

where $0 < \sigma_{\text{cool}} \leq 1$ denotes the cooling efficiency and CC_{CPUs} is calculated as $CC_{\text{CPUs}} = \sum_{j=1}^{J} CC_{\text{CPU},j}$. Moreover, $\Delta_{\text{GPP}}^{\text{proc}}$ stands for the slope of power consumption in a general purpose processor (GPP), and $CC_{\text{GPP}}^{\max}$ is the maximum processing capacity of each GPP in GOPS. The term $\chi_{\text{CPUs}}^{\text{proc}}$ is obtained as $\chi_{\text{CPUs}}^{\text{proc}} = P_{\text{GPP},0}^{\text{proc}} \sum_{j=1}^{J} W_j$, with W_j denoting the number of active GPPs in each CPU j,

[3] One can further decrease or enhance the values of b_{kl} and b_l. However, a deeper investigation regarding selecting the best values of b_{kl} and b_l is out of the scope of this paper.

TABLE I

NUMBER OF COMPLEX MULTIPLICATIONS REQUIRED FROM CPUS AND RUS TO PERFORM CHANNEL ESTIMATION AND GENERATE THE COMBINING VECTORS IN EACH COHERENCE BLOCK FOR DIFFERENT PRECODING SCHEMES

Scheme	Channel estimation		Combining vector computation	
P-RZF	$CC^{\text{est}}_{\text{CPU},j}$	$\sum_{k\in\mathcal{K}_j^{\text{all}}} \left(N\tau_p + N^2\right) L_{jk}$	$CC^{\text{comb}}_{\text{CPU},j}$	$\sum_{k\in\mathcal{K}_j^{\text{pri}}} \left[\frac{\lvert\mathcal{S}_k\rvert^2+\lvert\mathcal{S}_k\rvert}{2} NL_k + \lvert\mathcal{S}_k\rvert^2 + \lvert\mathcal{S}_k\rvert\, NL_k + \frac{\lvert\mathcal{S}_k\rvert^3-\lvert\mathcal{S}_k\rvert}{3}\right]$
P-MMSE	$CC^{\text{est}}_{\text{CPU},j}$	$\sum_{k\in\mathcal{K}_j^{\text{all}}} \left(N\tau_p + N^2\right) L_{jk}$	$CC^{\text{comb}}_{\text{CPU},j}$	$\sum_{k\in\mathcal{K}_j^{\text{pri}}} \left[\frac{(NL_k)^2+NL_k}{2} \lvert\mathcal{S}_k\rvert + (NL_k)^2 + \frac{(NL_k)^3-NL_k}{3}\right]$
LP-MMSE	$CC^{\text{est}}_{\text{RU},l}$	$\left(N\tau_p + N^2\right) K_l$	$CC^{\text{comb}}_{\text{RU},l}$	$\frac{1}{2}(N^2+N)K_l + N^2K_l + \frac{1}{3}(N^3-N)$
MR	$CC^{\text{est}}_{\text{RU},l}$	$\left(N\tau_p + N^2\right) K_l$	$CC^{\text{comb}}_{\text{RU},l}$	-

and $P^{\text{proc}}_{\text{GPP},0}$ representing the power consumed by each active GPP during idle mode. The term W_j can be given by $W_j = \lceil CC_{\text{CPU},j}/CC^{\max}_{\text{GPP}} \rceil$, with $\lceil . \rceil$ being the ceiling operation. It is worth mentioning that (23) extends the load dependent power consumption model proposed in [32] to a multi-CPU scenario.

IV. SCALABLE UC D-MMIMO SYSTEMS WITH RESTRICTED PROCESSING CAPACITY

In scalable D-mMIMO systems, the network complexity does not grow with the number of UEs since the number of UEs that each RU 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 RU 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 RUs [9]. That is, as L increases, the number of RUs connected to the UE k (L_k) can also increase, resulting in more processing complexity from the network, where $L_k = |\mathcal{M}_k|$. To circumvent this issue, we rely on a strategy where each UE can be associated only with a finite number of RUs, denoted as C_{max}, with $L_k \leq \mathrm{C}_{max}$ [19]. We refer to this strategy as maximum RU cluster size control. It is noteworthy that despite having a similar function, the C_{max} on this work is fundamentally different from the one presented in [19]. In this paper, C_{max} is a parameter that refers to the system processing capacity limitation that provides a new type of analysis for UC D-mMIMO systems.

A. RU Cluster Size Control With CPUs Cooperation

The maximum RU cluster size control procedure can be described as follows: when a new UE k enters the network, it measures the large-scale fading coefficients of the RUs in its vicinity, which is calculated according to $\beta_{kl} = \mathrm{tr}\left(\mathbf{R}_{kl}\right)/N$ [9]. Then, it claims a coordination RU to ensure its connection with at least one RU. The coordination RU serves the UE even if it has a poor channel condition [8]. The UE k points the RU l with

$$\begin{aligned} l \;&= \arg\max_l \beta_{kl} \\ &\text{s.t}\; |\mathcal{A}_l| < \tau_p, \end{aligned} \tag{24}$$

to be its coordination RU, where $\mathcal{A}_l \subset \mathcal{D}_l$ represents the subset of UEs the RU l serves as coordination RU. In order to solve (24), the UE k requests a connection to the available RUs. Posteriorly, the available RUs respond, and the UE k chooses the one with the strongest channel gain β_{kl} to be its coordination RU. The available RUs are the ones presenting $|\mathcal{A}_l| < \tau_p$, $\forall l \in \{1,\ldots,L\}$. Furthermore, $|\mathcal{B}_l| + |\mathcal{A}_l| \leq K_l$, where $\mathcal{B}_l \subset \mathcal{D}_l$ represents the subset of UEs the RU serves, but not as a coordination[4] (i.e., UEs that the RU may disconnect). It is worth mentioning that the CPU connected to the UE's coordination RU will be considered the UE's primary CPU [16].

After selecting the coordination RU, the UE k performs any UC RU selection scheme[5] in (6). In the following, the CPUs associated with the RU cluster of the UE k share the indexes of the RUs serving the UE ($\mathcal{M}_k$) with each other. Then, the CPUs serving the UE k compute the number of RUs 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 RUs presenting the weakest channel gains, where E_k denotes the number of RUs 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 RU cluster of the UE k. The maximum RU cluster size control is performed in J_k CPUs, where $J_k = |\mathcal{J}_k|$.

In order to drop the RUs in excess, the J_k CPUs serving the UE k sort the channel gains (β_{kl}) of the RUs serving the UE k in ascending order, such that $\tilde{\beta}_{kl'} \leq \cdots \leq \tilde{\beta}_{k(L_k)}$, where $\tilde{\beta}_{kl'}$ denotes the sorted version of β_{kl}, $\forall l \in \mathcal{M}_k$. The indexes of the RUs 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 RUs presenting the smallest channel gains after the sort operation. This procedure is given by

$$\mathrm{c}_{kl} = \begin{cases} 0 & \text{if } l' \leq E_k \\ 1 & \text{otherwise,} \end{cases} \tag{25}$$

where l' is mapped to the unsorted value of l in subset $\bar{\mathcal{M}}_k$. Hence, the final RU cluster of UE k will only be composed of the C_{max} RUs with the largest channel gains. Algorithm 1 summarizes the maximum RU cluster size control algorithm performed by the CPUs serving the UE.

The time complexity of the proposed method is computed as follows: the complexity for choosing a coordination RU by

[4]Subset $\mathcal{B}_l$ does not affect the coordination RU assignment in (24). For instance, if $K_l = \tau_p$ and $|\mathcal{B}_l| \geq 1$, the RU l could drop the UE with the weakest channel gain of $\mathcal{B}_l$ to serve the UE k in subset $\mathcal{A}_l$.

[5]The efficiency of the proposed solution is proportional to the effectiveness of the RU selection method. Thus, if the RU selection method does not provide a connection for a given UE, the proposed technique will not be activated.

solving (24) is $\mathcal{O}(L)$. The time complexity to perform RU cluster size control in each CPU j is $\mathcal{O}\big(\big|\mathcal{K}_j^{\text{all}}\big|\log\big|\mathcal{K}_j^{\text{all}}\big|\big)$, since each CPU has to perform a sort operation before computing (25). Thus, the overall time complexity can be expressed as $\mathcal{O}\big(L+\sum_{j=1}^{J}\big|\mathcal{K}_j^{\text{all}}\big|\log\big|\mathcal{K}_j^{\text{all}}\big|\big)$.

Remark 2: It is considered that controlling RU cluster size is done before reducing the effects of inter-communication between CPUs. In other words, we first limit the number of CPUs serving the UE so that $L_k \leq \mathrm{C}_{max}$. Then, we reduce the effects of inter-CPU communication in (14). This sequence also applies to the other methods in Sections IV and V. That is, they are performed before computing (14).

B. RU Cluster Size Control Without Using CPUs

This subsection presents a method that does not utilize the CPUs for controlling the maximum RU cluster sizes. Instead, it shows that RU selection methods themselves can incorporate C_{max}. Therefore, it is considered that the proposed approach is performed only between UEs and RUs. This one relies on a matched-decision (MD) strategy to make the RU clusters meet the restriction C_{max}. The MD strategy is a general RU selection framework that makes the RU clusters be composed of the more convenient connections for UEs and RUs. It is divided into two steps, where the UE first requests a connection to a subset of RUs following a decision criterion, e.g., largest-large-scale fading [5]. In the following, the RUs accept or reject the UE request using criteria such as least pilot contamination [9], effective channel gain [11], among others.

The subset of RUs selected by UE k is denoted by $\mathbf{e}_k = [\mathrm{e}_{k1},\ldots,\mathrm{e}_{kL}] \in \mathbb{N}^{1\times L}$, where $\mathrm{e}_{kl} = 1$ if the RU is in accordance with the criterion adopted by the UE. Otherwise, $\mathrm{e}_{kl}=0$. It is assumed that the UE k solves (24) and requests a connection to the C_{max} RUs presenting the strongest channel gains β_{kl} in its vicinity. For this, the UE k sorts the channel gains of the RUs in descending order, such that $\tilde{\beta}_{kl'} \geq \cdots \geq \tilde{\beta}_{kL}$, where $\tilde{\beta}_{kl'}$ denotes the sorted version of β_{kl}, for $l = \{1,\ldots,L\}$. The indexes of the RUs before the sort operation are stored in the l'-th element of the subset $\bar{\mathcal{M}}_k = \{1,\ldots,L\}$. Then, the UE requests a connection to the first C_{max} RUs presenting the strongest channel gains after the sort operation. This procedure can be expressed as

$$\mathrm{e}_{kl} = \begin{cases} 1 & \text{if } (l' \leq \mathrm{C}_{max}) \vee (\text{UE } k \in \mathcal{A}_l) \\ 0 & \text{otherwise,} \end{cases} \tag{26}$$

where l' is mapped to the unsorted value of l in subset $\bar{\mathcal{M}}_k$. The decision taken in the RU l is denoted by $\mathbf{f}_k = [\mathrm{f}_{k1},\ldots,\mathrm{f}_{kL}] \in \mathbb{N}^{1\times L}$, where $\mathrm{f}_{kl} = 1$ if the RU l accepts the UE k. Otherwise, $\mathrm{f}_{kl}=0$. It is considered that the RUs also rely on the channel gain to use similar decision criteria among the UEs and RUs. Therefore, f_{kl} is expressed as

$$\mathrm{f}_{kl} = \begin{cases} 1 & \text{if } (\text{UE } k \in \mathcal{A}_l) \vee (K_l < \tau_p) \vee \big(\beta_{kl} > \beta_{il}^{\min}\big) \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

where $i \neq k$ denotes the UE with the smallest channel gain that the RU l serves in $\mathcal{B}_l$. It is worth noting that in (27), the RU l drops the UE i to serve the UE k if $K_l = \tau_p$. Finally, the RU cluster of the UE k is given by

$$\mathbf{c}_k = \mathbf{e}_k \wedge \mathbf{f}_k. \tag{28}$$

Therefore, the connections will only be established if both UEs and RUs agree. Algorithm 2 summarizes the maximum RU cluster size control without using CPUs.

Algorithm 1 RU Cluster Size Control: CPUs

Input: $l = 1,\ldots,L$, C_{max}
The UE connects to a coordination RU by solving (24) and associates with a subset of RUs ($\mathcal{M}_k$) in (6);
Identify the J_k CPUs serving the UE; // $J_k = |\mathcal{J}_k|$
// The J_k CPUs perform RU 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 RUs 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 (25)
  end
end
Output: $\mathbf{c}_k = [\mathrm{c}_{k1},\ldots,\mathrm{c}_{kL}]$.

Algorithm 2 RU Cluster Size Control: RUs and UEs

Input: C_{max}, $\bar{\mathcal{M}}_k = \{1,\ldots,L\}$
The UE k connects to a coordination RU by solving (24) and sorts the channel gains (β_{kl}) of the RUs in descending order, such that $\tilde{\beta}_{kl'} \geq \cdots \geq \tilde{\beta}_{kL}$.
$\mathrm{e}_{kl} = 0$; $\mathrm{f}_{kl} = 0$, $\forall l \in \bar{\mathcal{M}}_k$.
for $l' = 1$ **to** C_{max} **do**
  Map l' to the unsorted value of l in subset $\bar{\mathcal{M}}_k$.
  $\mathrm{e}_{kl} = 1$; // Request a connection to nearby RUs
  // RUs 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 $K_l = \tau_p$ **then**
      $\mathrm{f}_{il} = 0$;
    end
  end
  $\mathrm{c}_{kl} = (\mathrm{e}_{kl} \wedge \mathrm{f}_{kl})$ // Matched-decision
end
Output: $\mathbf{c}_k = [\mathrm{c}_{k1},\ldots,\mathrm{c}_{kL}]$.

V. RU Cluster Adjustment

In this section, a heuristic method that adjusts the RU clusters according to the network implementation is proposed. Such method holds for any UC RU selection scheme, i.e., with and without processing capacity limitation. Besides, it is a heuristic strategy because only heuristic solutions are scalable [9]. In a nutshell, the UEs are associated with a subset of RUs following any RU selection process. Then, the proposed method aims to simultaneously reduce the number

of UEs served by each RU l (K_l) and the number of RUs 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 D-mMIMO systems. Throughout the analysis, it is also assumed that each UE connects to a coordination RU.

A. RU Cluster Adjustment in the Distributed Implementation

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

The adjustment of the RU 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 use these metrics to prevent the less fortunate UEs from being easily dropped by the RU. Thus, they do not directly represent the long-term CSI of the UEs that the RU 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 RU, such that $0 < \lambda_l < 1$. Without this normalization, the RU could easily drop a UE presenting a weaker channel gain if the RU was also serving UEs with stronger channel gains. However, these differences can be reduced 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 $\lambda_l = \min_{k\in\mathcal{D}_l}(\beta_{kl})/\max_{k\in\mathcal{D}_l}(\beta_{kl})$. 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 RU l to calculate $\bar{\beta}_{l,-k} = \bar{\beta}_l - \bar{\beta}_{kl}$, $\forall k \in \mathcal{D}_l$. 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 RU l. After computing $\bar{\beta}_{l,-k}$, the RU keeps the connection of UE k, only if

$$\mathrm{c}_{kl} = \begin{cases} 1 & \text{if } (\text{UE } k \in \mathcal{A}_l) \vee \left(\bar{\beta}_{l,-k} \leq \bar{\beta}_l^{mean}\right) \\ 0 & \text{otherwise,} \end{cases} \tag{29}$$

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 RU l serves as a coordination. 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 RU l.

B. RU Cluster Adjustment in the Centralized Implementation

In the centralized implementation, the long-term CSI of RUs and UEs is available at the CPUs [7], [9]. 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 RUs serving the UE (recall that $L_k = |\mathcal{M}_k|$), the RU cluster expansion will not always be beneficial, and reducing L_k may be necessary even in this implementation. In the centralized implementation, the RU cluster adjustment is also performed by the J_k CPUs associated with the RU 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 RUs serving the UE ($\mathcal{M}_k$) with each other, as in Section IV.

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 RUs presenting the smallest and highest channel gain to the UE k. The CPUs compute λ_k as $\lambda_k = \min_{l\in\mathcal{M}_k}(\beta_{kl})/\max_{l\in\mathcal{M}_k}(\beta_{kl})$. 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 RU brings to $\bar{\beta}_k$ as $\bar{\beta}_{k,-l} = \bar{\beta}_k - \bar{\beta}_{kl}$, $\forall l \in \mathcal{M}_k$. Therefore, the CPU connected to the RU l keeps the connection of RU l with the UE k only if

$$\mathrm{c}_{kl} = \begin{cases} 1 & \text{if } (\text{UE } k \in \mathcal{A}_l) \vee \left(\bar{\beta}_{k,-l} \leq \bar{\beta}_k^{mean}\right) \\ 0 & \text{otherwise,} \end{cases} \tag{30}$$

where $\bar{\beta}_k^{mean} = \sigma_{si}/2 + \sum_{l\in\mathcal{M}_k}\bar{\beta}_{k,-l}/L_k$ and σ_{si} denotes 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 RUs from the RU 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 RU cluster of the UE run the proposed method.

C. Pros and Cons of the Two RU Clusters Adjustments

The utilization of the proposed method in a distributed implementation enables a fronthaul signaling reduction since the number of data flows on the fronthaul is proportional to K_l in (17). Besides, it allows the RU 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 required bit rate is not proportional to K_l in (17). It is worth noting that this paper has considered that the RU cluster adjustment is only activated when λ_l and λ_k are lesser than a threshold Γ to avoid excessive adjustments, where Γ is a project parameter. We have set $\Gamma = 10^{-2}$ and $\Gamma = 10^{-3}$ for the distributed and centralized implementations, respectively.

VI. Numerical Results

We consider a D-mMIMO network consisting of K single antenna UEs and L RUs, each equipped with N antennas. The values of L, N, and K vary and are specified throughout the results. The K UEs are uniformly distributed over a square area of 1×1 km, and the distribution of the RUs follows a hard

core point process (HCPP).[6] After the RUs positioning, the coverage area is divided into J rectangle regions of the same size, each consisting of a CPU coordinating approximately L/J RUs, where $J = 4$. Each CPU can serve up to $K^{\text{sec}} = 10$ UEs as a secondary CPU [16]. The simulations focus on DL channels and it is assumed that $\tau_c = 200$, $\tau_p = 10$, and $\tau_d = 190$. The total transmission powers of the UEs and RUs are 100 mW and 200 mW, respectively. We perform Monte-Carlo simulations to account for different RU/UE locations and channel realizations. The wrap-around technique is also utilized to provide a better balance regarding the amount of interference affecting each RU.

We utilize an RU clustering scheme that jointly performs the pilot assignment and RU selection [8]. In this one, the UEs can connect to coordination and non-coordinations RUs. The non-coordinations serve only the UEs with the greatest channel gain in each pilot. The first τ_p UEs are assigned to mutually orthogonal pilots, and the remaining ones to the pilot causing the lowest pilot contamination. Hereafter, we name it as scalable cell-free (SCF) scheme. Furthermore, we utilize two other RU selection strategies to assess the proposed approach's performance in UC D-mMIMO systems presenting distinct RU-UE association strategies. Both are non-scalable solutions but are utilized to demonstrate that the proposed approach can be used in any RU selection scheme when RU cluster size control is performed at the CPUs. The key features of these RU selection methods are described below:

- Largest-large-scale fading (LSFB) [5]: the UE measures the large-scale fading gains of the RUs in its surroundings and sums these channel gains. Posteriorly, it connects to the subset of RUs that contribute the most to the sum of its total channel gain in percentage $\delta\%$, with $\delta\% = 99.9$. The LSFB is a non-scalable scheme because it does not limit the number of UEs that each RU can serve.
- User-centric clustering (UCC) [6]: the RU serves the $U_{\max}$ UEs with the largest estimated channel in each coherence block, where $U_{\max}$ is the maximum number of UEs the RU can serve. We have adjusted this strategy to consider only the large-scale fading to avoid performing RU selection in each coherence block. Thus, the RU serves the $U_{\max}$ UEs presenting the largest large-scale fading in their vicinity, with $U_{\max} = \tau_p$.

The 3GPP Urban Micro (UMi) path loss model is adopted for modeling the propagation channel, with LOS/NLOS conditions defined in the Technical Report (TR) 38.901 [33]. It is considered that the shadowing terms of an RU to different UEs are correlated, and the computation of correlation matrices $\mathbf{R}_{kl}$ follows the local scattering spatial correlation model [8]. Table II exhibits the parameters used in the UMi and $\mathbf{R}_{kl}$ models [8], [34].

The power coefficients at RU l in the distributed implementation are set as $\rho_{kl} = \rho_d \sqrt{\beta_{kl}} / \sum_{k' \in \mathcal{D}_l} \sqrt{\beta_{k'l}}$, where ρ_d is the maximum transmit power per RU. For the centralized one, scalable fractional power control is used with the following parameters: $v = -0.5$ and $\kappa = 0.5$ [8]. The EE parameters related to the power consumption of the hardware of the RUs, fronthaul, and backhaul links are summarized in Table III, which follows [5], [29]. However, $P^{\text{proc}}_{\text{RU},0}$, $\Delta^{\text{proc}}_{\text{RU},l}$, and $CC^{\max}_{\text{RU}}$ are in accordance with a Texas Instruments TMS320C6678 DSP. Moreover, conventional 5G new radio (NR) parameters are assumed to compute the CC in GOPS, where $N_{DFT} = N_{sc} = 3300$, $f_s = 122.88$MHz, and $T_s = 35.38\mu$s. These values correspond to 30kHz of subcarrier spacing.

[6]This method states that the distance between any two RUs cannot be smaller than $d_{\min} = \sqrt{A/L}$, where A is the coverage area in square meters. The first step is to randomly drop the RUs based on a homogeneous Poisson point process with mean rate $1/d_{\min}$, then randomly update the location of RUs that do not meet the spacing requirement until it is fulfilled.

TABLE II
PARAMETERS ASSUMED FOR THE UMI PATH LOSS AND LOCAL SCATTERING SPATIAL CORRELATION MODEL

Parameter	Value
Shadow fading standard deviation, σ_{SF}	4 dB
RU/UE antenna height, $h_{\text{RU}}, h_{\text{UE}}$	11.65 m, 1.65 m
RX noise figure (NF)	8 dB
Carrier frequency, bandwidth (B)	3.5GHz, 100MHz
Angular standard deviations (ASDs)	$\sigma_\varphi = \sigma_\theta = 15^\circ$
Antenna spacing	1/2 wavelength distance

TABLE III
PARAMETERS ASSUMED FOR CALCULATING THE POWER CONSUMPTION IN CPUS, BACKHAUL/FRONTHAUL LINKS, AND EE

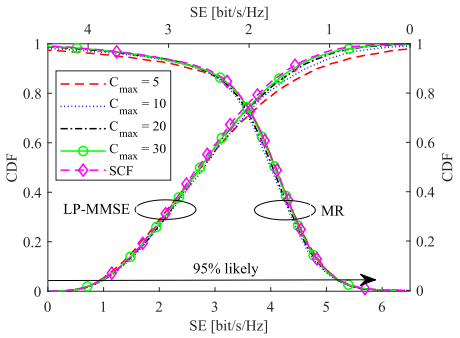
Parameter	Value	Parameter	Value
$P^{\text{proc}}_{\text{RU},0}$, $P^{\text{proc}}_{\text{GPP},0}$	7.3 W, 20.8 W	$P_{0,l}$, $P_{\text{bh},0}$	0.825 W
$\Delta^{\text{proc}}_{\text{RU},l}$, $\Delta^{\text{proc}}_{\text{GPP}}$	73 mW, 74 W	$P_{\text{ft},l}$, $P_{\text{bt},l}$	0.25 W/(Gbit/s)
$CC^{\max}_{\text{RU}}$, $CC^{\max}_{\text{GPP}}$	180 GOPS	σ_{cool}, γ_l	0.9, 0.4
$R^{\max}_{\text{fh},l}$	10 Gbps	$b^{\max}$	12 bits

A. Impacts of Limiting the Processing Capacity

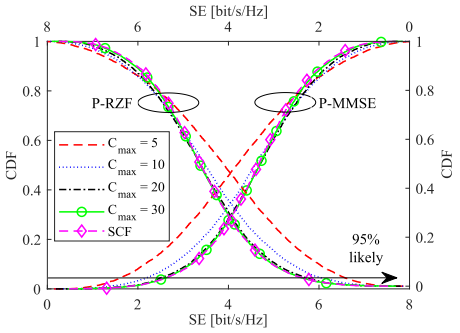
We start by evaluating a network composed of $K = 25$ UEs and $L = 100$ RUs equipped with $N = 1$ antenna. Fig. 2 presents the cumulative distribution functions (CDFs) of the SE of UC systems with and without processing capacity limitation. It considers different processing capacity limitations, i.e., several values of C_{max}, and the system is compared with a traditional UC scheme (i.e., L_k and K^{sec} are not restricted), which we have denoted as SCF. These results are presented for the case where CPUs impose that $L_k \leq \text{C}_{max}$.

In Fig. 2a, the SE is not significantly reduced by the variations of C_{max}. The SE even increases slightly for $5 \leq \text{C}_{max} \leq 10$. This is because decreasing L_k also reduces K_l, helping precoding techniques such as LP-MMSE (of local processing) to mitigate interference. In Fig. 2b, the SE can suffer significant losses when C_{max} is small. For instance, it decreases by 32% when C_{max} goes from 30 to 5. Hence, reducing the RU cluster sizes (L_k) may lead the centralized implementation to not exploit its full potential in mitigating interference and improving SE. Therefore, this implementation needs to utilize more processing capacity, such as $\text{C}_{max} \geq 20$.

Fig. 3 presents the SE and sum of CC to perform channel estimation and generate the combining vectors when the number of RUs L varies. The CC is given in terms of number

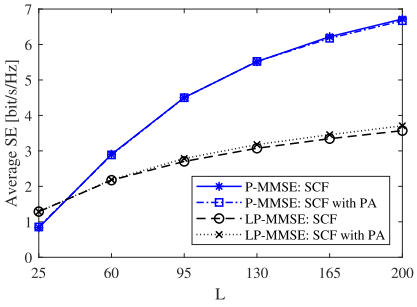


(a) Distributed implementation.

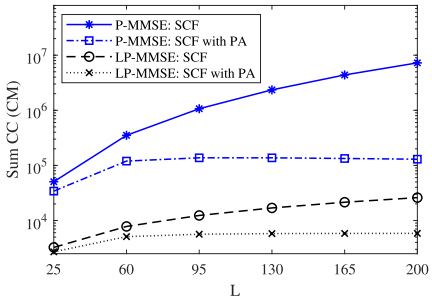


(b) Centralized implementation.

Fig. 2. CDF of SE by varying C_{max} from 5 to 30. Parameters setting: $J = 4$, $L = 100$, $K = 25$, $N = 1$, and $K^{\text{sec}} = 10$.



(a) SE



(b) CC

Fig. 3. Average DL SE (a) and CC in terms of number of complex multiplications (CM) (b) achieved by varying the number of RUs L. Parameters setting: $J = 4$, $K = 25$, $N = 1$, $\mathrm{C}_{max} = 20$, and $K^{\text{sec}} = 10$.

of complex multiplications (CM), i.e., only considering the terms $CC^{\text{cecb}}_{\text{CPU},j}$ and $CC^{\text{cecb}}_{\text{RU},j}$ in (19) and (20). In Fig. 3a, the average SE grows with L for UC systems with and without the proposed approach for processing capacity limitation. Despite this, limited systems have a significant advantage, as their CC does not always increase with L, starting to be constant from $L = 60$. This behavior occurs because K_l and L_k does not increase with L, as Table IV demonstrates. Additionally, it is possible to observe that the CC decreases by about 98.22% when the proposed approach for processing capacity limitation is employed together with the P-MMSE scheme for $L = 200$. It is noteworthy that L_k does not equal C_{max} in Table IV. This happens because after applying (28), the proposed approach (PA) also reduces the effects of inter-CPU communication in (14). Thus, L_k is lowered both in (28) and (14), making $L_k < \mathrm{C}_{max}$.

Fig. 4 shows the total CC in GOPS when the number of RUs L varies, i.e., considering all the terms of (19) and (20). One can note that the overall CC still increases with the number of RUs. This is because some network functions, such as higher-layer control, do not rely on C_{max}. Therefore, their CC (denoted as $CC^{\text{basic}}_{\text{CPU},j}$) still grows with the number of RUs L. However, the proposed approach still effectively reduces the CC. For instance, it decreases the CC by about 85% and 77.5% in centralized and distributed implementations, respectively.

Fig. 5 presents the backhaul traffic for UC D-mMIMO systems, with and without the proposed approach, when the number of UEs K grows. It can be noted that the proposed approach (i.e., using C_{max} and K^{sec}) decreases the backhaul traffic significantly. For instance, it reduces the backhaul traffic by about 77% and 80% in centralized and distributed implementations, respectively, for $K = 100$. These results indicate that the proposed strategy allows UC D-mMIMO systems to reduce their CC and signaling demands while keeping the SE under minor degradation.

Fig. 6 illustrates the EE achieved by a UC system with and without the proposed approach for processing capacity limitations with different values of C_{max}. Note that the processing capacity limitation can provide a considerable improvement in the EE, especially for small values of C_{max}. For instance, the EE grows by about 47% in the LP-MMSE and 106% in the P-MMSE, when C_{max} decreases from 50 to 5. This happens because reducing C_{max} also decreases the number of UEs the RUs serve (i.e., K_l), as indicated in Table IV. For instance, the proposed approach reduces the average value of K_l from 10 to 2.25 for $L =$

TABLE IV

AVERAGE NUMBER OF RUS PER UE (L_k) AND UES PER RU (K_l) WITHOUT AND WITH RU CLUSTER CONTROL. PARAMETERS SETTING: $J = 4$, $K = 25$, $N = 1$, $\mathrm{C}_{max} = 20$, AND $K^{\text{sec}} = 10$

Method	$L = 95$		$L = 200$	
	K_l	L_k	K_l	L_k
SCF	10	38	10	80
With PA	4.56	17.35	2.25	18

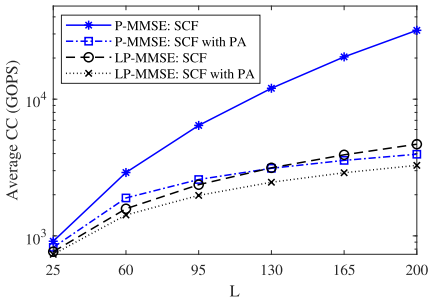


Fig. 4. Average CC in GOPS achieved by varying the number of RUs L. Parameters setting: $J = 4$, $K = 25$, $N = 1$, $\mathrm{C}_{max} = 20$, and $K^{\mathrm{sec}} = 10$.

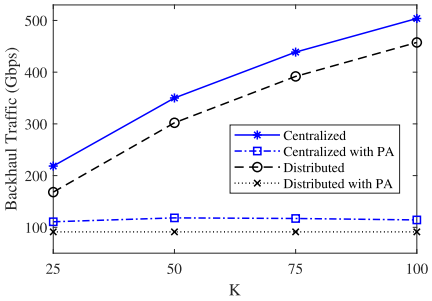


Fig. 5. Backhaul traffic in each network implementation by varying the number of UEs K. Parameters setting: $J = 4$, $L = 100$, $N = 1$, $\mathrm{C}_{max} = 20$, and $K^{\mathrm{sec}} = 10$.

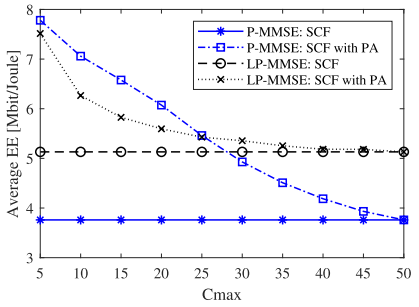


Fig. 6. Average EE achieved by varying C_{max}. Parameters setting: $J = 4$, $L = 100$, $K = 25$, $N = 1$, and $K^{\mathrm{sec}} = 10$.

200. Consequently, the power consumption in each fronthaul link $P_{\mathrm{fh},l}$ reduces since $P_{\mathrm{fh},l}$ is proportional to the number of UEs served by the RU. Thus, even though the system experiences SE losses when $\mathrm{C}_{max} = 5$, the reduction in power consumption in the fronthaul links compensates for these losses, thereby increasing the EE. In other words, the denominator of (21), which contains $P_{\mathrm{fh},l}$, decreases more than the numerator, which contains the SE.

Fig. 7 compares the RU cluster size control performed by the CPUs with the MD scheme, which operates between UEs and RUs. One can note that both methods maintain the SE with minor degradation compared to the SCF scheme. It is noteworthy that CPUs provide adaptable control for any RU

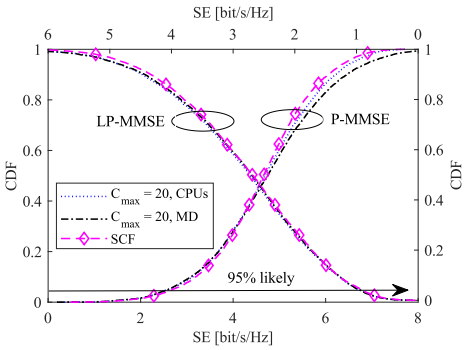


Fig. 7. CDF of SE for the proposed strategies to control the RU cluster size. Parameters setting: $J = 4$, $L = 100$, $K = 25$, $N = 1$, and $K^{\mathrm{sec}} = 10$.

selection method, while the MD strategy has limited flexibility as it is performed only between UEs and RUs. However, the MD scheme demonstrates that the benefits of limiting CC can be achieved whether the procedure of computing and applying $\mathrm{C}_{\max}$ is centralized in CPUs or occurs locally among the UEs and RUs.

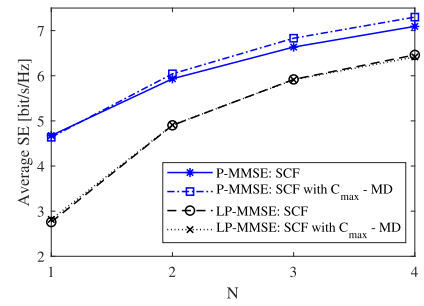
Fig. 8 presents the average SE and the total CC in GOPS as the number of antenna elements per RUs N varies. The MD scheme is used to control the maximum RU cluster size. As expected, the proposed approach reduces the overall CC in GOPS, as it mitigates the impact of N, as shown in Table I. Interestingly, the proposed approach also leads to some improvements in SE as N increases in the P-MMSE scheme. This happens because when the RU cluster size for certain UEs is large, these UEs may connect to RUs that contribute marginally to the desired signal while intensifying interference. Consequently, removing such RUs can be beneficial, particularly in sophisticated precoding techniques like P-MMSE, as these techniques can better manage interference and further exploit the advantages of using more antenna elements per RU when the UE is connected to fewer RUs.

However, this same effect is not observed in the LP-MMSE scheme since its interference mitigation is performed locally at the RUs, which do not have access to global statistical and channel estimation information. Additionally, although not shown in the figures due to space constraints, our results demonstrate that the same behaviors are observed when the RU cluster size control is performed at the CPUs.

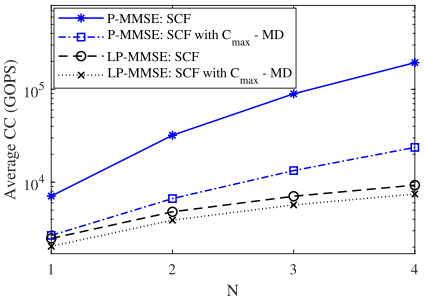
B. Comparison With Other Baseline Solutions

Recall that the proposed approach can be applied to any RU selection strategy when RU cluster size control is performed at the CPUs, as described in Subsection IV-A. Therefore, we also evaluate the proposed approach for the UCC and LSFB RU selection schemes, which are non-scalable strategies but can be utilized to demonstrate the proposed approach's performance. To avoid redundancies, we present the results only for the P-MMSE precoding scheme.

In this regard, Fig. 9 shows the average SE and CC in GOPS as the number of RUs L, varies. It can be noted that the previous explanations hold even if the RU selection scheme

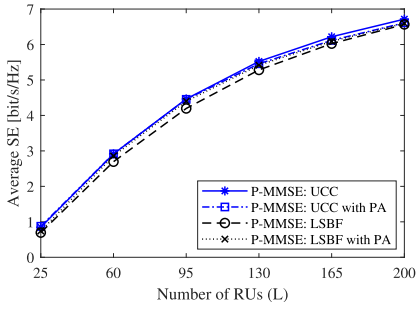


(a) SE

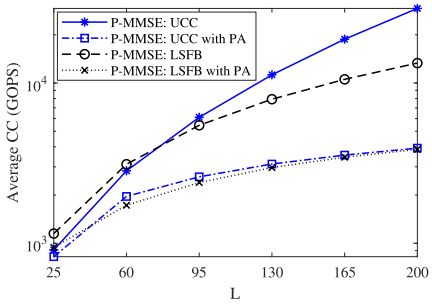


(b) CC

Fig. 8. Averages DL SE (a) and CC in GOPS (b) achieved by varying the number antennas per RU N, when the MD scheme is utilized to control the RU cluster size. Parameters setting: $J = 4$, $K = 25$, $L = 100$, $C_{max} = 20$, and $K^{\text{sec}} = 10$.



(a) SE



(b) CC

Fig. 9. Averages DL SE (a) and CC in GOPS (b) achieved by varying the number of RUs L, when the UCC and LSFB RU selection schemes are employed. Parameters setting: $J = 4$, $K = 25$, $N = 1$, $C_{max} = 20$, and $K^{\text{sec}} = 10$.

is modified, i.e., the proposed approach can reduce the CC while yielding minor degradation in the SE. Besides, it is possible to observe that distinct RU selection schemes can lead to different CC and SE since they operate differently and thus affect the overall CC distinctly. For instance, the proposed approach decreases the average CC by about 86% and 54.5% for the UCC and LSFB schemes, respectively. It is worth mentioning that the higher CC of the UCC scheme compared to the LSFB is associated with the fact that the UCC scheme always makes the RUs operate at their maximum capacity, i.e., $K_l = \tau_p$. In contrast, K_l is usually lesser than τ_p in the LSFB strategy.

The proposed approach can also improve the EE for both the UCC and LSFB schemes, as depicted in Fig. 10. The EE increases by about 56% and 94% in the LSFB and UCC schemes, respectively, when C_{max} decreases from 50 to 5. The higher EEs of the LSFB scheme is also related to the fact that it operates with RUs serving fewer UEs, which reduces not only the CC but also the power consumption in the fronthaul links, thereby improving the EE.

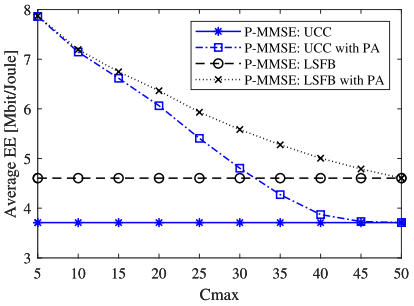


Fig. 10. Average EE achieved by varying C_{max}, when the UCC and LSFB RU selection schemes are employed. Parameters setting: $J = 4$, $L = 100$, $K = 25$, $N = 1$, and $K^{\text{sec}} = 10$.

TABLE V

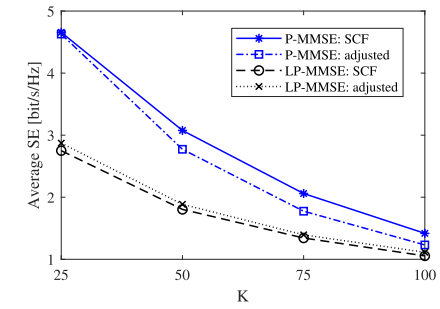
AVERAGE NUMBER OF RUS PER UE (L_k) AND UES PER RU (K_l) WITHOUT AND WITH RU CLUSTER ADJUSTMENT. PARAMETERS SETTING: $L = 100$, $N = 1$, AND $K^{\text{sec}} = 10$

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

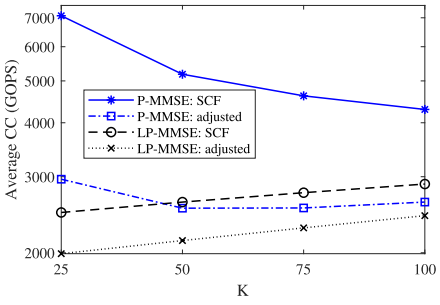
C. Impacts of RU Cluster Adjustment

From now on, we will investigate the impacts of adjusting the RU clusters in UC D-mMIMO systems. We will focus on UC systems without processing capacity limitation to assess the full benefits of the RU 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.

Fig. 11 presents the average SE and total CC versus the number of UEs K in a network composed of $L = 100$ RUs



(a) SE



(b) CC

Fig. 11. Average DL SE (a) and CC in GOPS (b) achieved by varying the number of UEs K, when the proposed RU cluster adjustment is employed. Parameters setting: $J = 4$, $L = 100$, $N = 1$, and $K^{\text{sec}} = 10$.

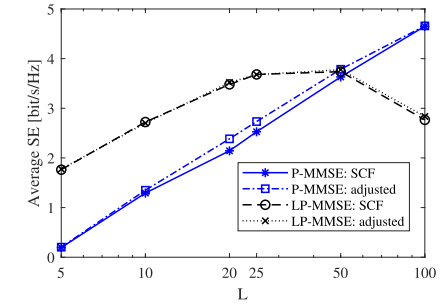


Fig. 12. Average DL SE by varying L and N, while keeping $M = LN = 100$, when the proposed RU cluster adjustment is employed. Parameters setting: $J = 4$, $K = 25$ and $K^{\text{sec}} = 10$.

equipped with $N = 1$ antenna. Note that the SE is significantly reduced in the P-MMSE with the proposed method (denoted as adjusted). Nonetheless, this reduction is related to the value of $K^{\text{sec}} = 10$, which is small for centralized implementation. For instance, although not shown in the figures due to space constraints, the average SE for $K^{\text{sec}} = 20$ has been analyzed, and the RU cluster adjustment revealed to affect the SE negligibly. One can also note that the proposed method causes a slight increase in the SE of LP-MMSE. Moreover, the RU cluster adjustment also reduces the CC of both network implementations, decreasing the CC by up to 58% in the P-MMSE scheme for $K = 25$. Finally, the proposed method decreases the values of K_l and L_k as illustrated in Table V, indicating that it can also increase the EE.

Fig. 12 presents the average SE and CC versus the number of RUs L and N for a fixed total number of antennas $M = LN = 100$, when $K = 25$. One can note that the same discussions regarding decreasing CC apply to this case. The difference is the SE behavior. When $L = 50$ and $N = 2$, 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 RU clusters are adjusted for $10 < L < 100$. This is because the fewer RUs in the coverage area, the further away the RUs will be from the UE. Therefore, disconnecting some of these RUs will not impact the UE's performance [12].

VII. Conclusion

This paper investigated the performance of scalable UC D-mMIMO systems with multiple CPUs. In this regard, we proposed a framework that restricts the number of RUs serving each UE and considers that each CPU can only serve a limited number of UEs managed by other CPUs. The backhaul traffic was modeled, and the CC of performing channel estimation and generating the combining vectors was presented for multiple CPUs in four precoding schemes. The EE modeling was also improved by considering the power consumption in CPUs and backhaul links. Two strategies for controlling the size of the RU cluster were presented, where one was carried out by the CPUs and the other between the UEs and the RUs. Moreover, we proposed two methods that adjust the RU clusters to the network implementations, i.e., centralized and distributed. The results demonstrated that using the proposed strategies to restrict the processing capacity can improve the EE up to 106% in centralized implementation. However, it can degrade the SE of centralized implementation when the maximum number of RUs serving the UE is small. On the other hand, RU clusters comprising just a few RUs almost do not harm the SE of distributed implementation. The benefits of the proposed schemes were observed in both RU cluster size control methods presented in this paper.

Simulation results also reveal that the proposed framework allows UC D-mMIMO systems to decrease CC and signaling requirements while maintaining minor degradation in SE. For instance, it can reduce the CC to perform channel estimation and generate the combining vectors up to 98% while preventing it from growing with the number of RUs. The backhaul traffic due to inter-CPU communication is also controlled, i.e., it does not increase with the number of UEs. Nonetheless, the CC of certain network functions, such as higher-layer control, are not affected by the proposed methods. Thus, their CC continue to scale with the number of RUs. Finally, the proposed RU cluster adjustment can slightly improve the SE of distributed implementation while reducing the CC in both network implementations. These results open the way for future works to design practical UC systems with processing capacity and signaling constraints. Future works can expand our analyses to consider non-reciprocity and mobility.

REFERENCES

[1] M. M. M. Freitas et al., "Scalable user-centric distributed massive MIMO systems with limited processing capacity," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 4298–4304.

[2] 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. 99878–99888, 2019.

[3] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133995–134030, 2020.

[4] 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.

[5] H. Q. Ngo, L.-N. 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, Mar. 2018.

[6] 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.

[7] 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.

[8] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends Signal Process.*, vol. 14, pp. 162–472, Jan. 2021.

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

[10] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[11] H. T. Dao and S. Kim, "Effective channel gain-based access point selection in cell-free massive MIMO systems," *IEEE Access*, vol. 8, pp. 108127–108132, 2020.

[12] M. Freitas et al., "Matched-decision AP selection for user-centric cell-free massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6375–6391, May 2023.

[13] 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. 14462–14473, 2018.

[14] 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. (GLOBECOM)*, Dec. 2021, pp. 1–6.

[15] V. Ranjbar, A. Girycki, M. A. Rahman, S. Pollin, M. Moonen, and E. Vinogradov, "Cell-free mMIMO support in the O-RAN architecture: A PHY layer perspective for 5G and beyond networks," *IEEE Commun. Standards Mag.*, vol. 6, no. 1, pp. 28–34, Mar. 2022.

[16] M. M. M. Freitas et al., "Reducing inter-CPU coordination in user-centric distributed massive MIMO networks," *IEEE Wireless Commun. Lett.*, vol. 12, no. 6, pp. 957–961, Jun. 2023.

[17] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "Distributed resource allocation optimization for user-centric cell-free MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3099–3115, May 2022.

[18] F. Li, Q. Sun, X. Ji, and X. Chen, "Scalable cell-free massive MIMO with multiple CPUs," *Mathematics*, vol. 10, no. 11, p. 1900, Jun. 2022.

[19] D. D. Souza et al., "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.

[20] Ö. 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.

[21] Ö. 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, Nov. 2019.

[22] A. Á. Polegre, F. Riera-Palou, G. Femenias, and A. G. Armada, "Channel hardening in cell-free and user-centric massive MIMO networks with spatially correlated ricean fading," *IEEE Access*, vol. 8, pp. 139827–139845, 2020.

[23] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.

[24] A. H. Jafari, D. Lopez-Perez, M. Ding, and J. Zhang, "Study on scheduling techniques for ultra dense small cell networks," in *Proc. IEEE 82nd Veh. Technol. Conf.*, Sep. 2015, pp. 1–6.

[25] G. Femenias and F. Riera-Palou, "Fronthaul-constrained cell-free massive MIMO with low resolution ADCs," *IEEE Access*, vol. 8, pp. 116195–116215, 2020.

[26] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inf. Theory*, vol. IT-6, no. 1, pp. 7–12, Mar. 1960.

[27] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, vol. 159. Berlin, Germany: Springer, 2012.

[28] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–7.

[29] Ö. T. Demir, M. Masoudi, E. Björnson, and C. Cavdar, "Cell-free massive MIMO in O-RAN: Energy-aware joint orchestration of cloud, fronthaul, and radio resources," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 356–372, Feb. 2024.

[30] S. Malkowsky et al., "The world's first real-time testbed for massive MIMO: Design, implementation, and validation," *IEEE Access*, vol. 5, pp. 9073–9088, 2017.

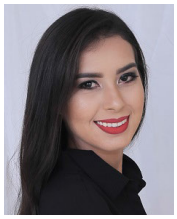
[31] C. Desset and B. Debaillie, "Massive MIMO for energy-efficient communications," in *Proc. 46th Eur. Microw. Conf. (EuMC)*, Oct. 2016, pp. 138–141.

[32] T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, "Energy-efficient cloud radio access networks by cloud based workload consolidation for 5G," *J. Netw. Comput. Appl.*, vol. 78, pp. 1–8, Jan. 2017.

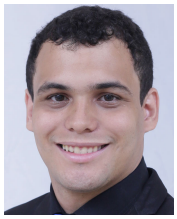
[33] *Study on Channel Model for Frequencies From 0.5 To 100 GHz*, document TR 38.901 (Release 16), 3GPP, 2019.

[34] *NR; User Equipment (UE) Radio Transmission and Reception; Part 1: Range 1 Standalone*, document TR 38.101-1 (Release 17), 3GPP, 2021.

Marx M. M. Freitas received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Federal University of Pará (UFPA), Belém, Brazil, in 2018, 2019, and 2024, respectively. From 2022 to 2023, he was a Visiting Ph.D. Student with the Scuola Superiore Sant'Anna, Pisa, Italy. He has experience in broadband communication systems. His research interests include signal processing, resource allocation, massive MIMO, and mobile transport networks for future wireless communication systems.



Daynara D. Souza received the B.Sc. degree in electrical engineering and the M.Sc. degree in electrical engineering with an emphasis on telecommunication from the Federal University of Pará (UFPA), Belém, Brazil, in 2018 and 2020, respectively. She is currently pursuing the Ph.D. degree with UFPA in partnership with the School of Energy Systems, LUT University, Finland. Her research interests include signal processing, resource allocation, and the development of emergency communication networks utilizing cell-free massive MIMO systems.



André L. P. Fernandes received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Federal University of Pará, Belém, Brazil, in 2018, 2019, and 2024, respectively. He was a Visiting Researcher with the Chalmers University of Technology, Gothenburg, Sweden, from 2022 to 2023. He has experience in techno-economic modeling for fixed and mobile broadband communication systems and reliability analysis for communication systems. His research interests include enabling technologies for future wireless communication systems, such as massive MIMO, mobile transport infrastructure, and network deployment analysis and optimization.

Daniel Benevides da Costa (Senior Member, IEEE) received the B.Sc. degree in telecommunications from the Military Institute of Engineering (IME), Rio de Janeiro, Brazil, in 2003, and the M.Sc. and Ph.D. degrees in electrical engineering and telecommunications from the University of Campinas, Brazil, in 2006 and 2008, respectively. He is currently a Distinguished University Professor with the Department of Electrical Engineering, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. He was a recipient of six conference paper awards. His Ph.D. thesis received the Best Ph.D. Thesis in Electrical Engineering by the Brazilian Ministry of Education (CAPES) at the 2009 CAPES Thesis Contest. He has been recognized as World's Top 2% Scientist by Stanford University (2021–2024) and has been ranked among 1% Top Scientists in the world in the broad field of electronics and electrical engineering in 2022 and 2023. He is the Editor-in-Chief of IEEE COMMUNICATIONS LETTERS and a Specialty Chief Editor of *Frontiers in Communications and Networks*–Wireless Communications Section. He is a Distinguished Speaker of the IEEE Vehicular Technology Society.

André Mendes Cavalcante (Member, IEEE) received the Ph.D. degree in electrical engineering from the Federal University of Pará (UFPA), Brazil, in 2007. From 2007 to 2016, he was a Co-Worker with the Nokia Institute of Technology (INDT), Brazil, where he was involved in several research and development projects related to wireless communications. He is currently a Senior Researcher with Ericsson Research, Brazil. His research interests include distributed massive MIMO for future wireless communication systems.

Luca Valcarenghi (Senior Member, IEEE) received the Laurea degree in electrical engineering from the Politecnico di Torino in 1997 and the M.S.E.E. and Ph.D. degrees in electrical engineering (major in telecommunications) from UTD in 1999 and 2001, respectively. He has been a Full Professor with the Scuola Superiore Sant'Anna, Pisa, Italy, since 2014. He has published more than 300 papers in international journals and conference proceedings. He received the Fulbright Research Scholar Fellowship in 2009 and the JSPS "Invitation Fellowship Program for Research in Japan (Long Term)" in 2013. His main research interests are optical networks design, analysis, and optimization; communication networks reliability; energy efficiency in communications networks; optical access networks; zero touch network and service management; and 5G technologies and beyond.

João C. Weyl Albuquerque Costa (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Federal University of Pará (UFPA), Belém, Brazil, in 1981, the M.Sc. degree in electrical engineering from the Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, in 1989, and the Ph.D. degree in electrical engineering from the State University of Campinas, Campinas, Brazil, in 1994. He is currently a Professor with the Institute of Technology, UFPA; and a Researcher with the Brazilian Research Funding Agency National Council for Scientific and Technological Development, Brasília, Brazil. His current research interests include broadband systems and optical sensors.