Matched-Decision AP Selection for User-Centric Cell-Free Massive MIMO Networks

Marx Freitas, Daynara Souza, Daniel Benevides da Costa, *Senior Member, IEEE,* Gilvan Borges, André Mendes Cavalcante, *Member, IEEE,* Maria Marquezini, *Member, IEEE,* Igor Almeida, *Member, IEEE,* Roberto Rodrigues, and João C. W. A. Costa, *Senior Member, IEEE.*

***Abstract*—This work proposes a scalable access point (AP) selection framework based on a competitive mechanism that operates in two stages. Initially, the user equipment (UE) connects to an intermediate AP cluster, where a matched-decision among the APs and UEs establishes the best connection in terms of large-scale fading and prevents UEs from being dropped. Then, UEs can expand their AP clusters by connecting to more APs. We compare our method with baseline solutions considering different precoding techniques in centralized and distributed network implementations. We propose three strategies to fine-tune the AP clusters, reducing the number of UEs per AP without compromising the spectral efficiency (SE) or improving energy efficiency. The simulations comprise a range of UEs and APs, the number of antennas per AP varies, and each AP can serve a limited number of UEs. The results show that our solution improves up to 163% the SE of the 95% likely UEs compared with baseline solutions and that the number of UEs each AP serves is crucial for improving SE. Additional results indicate that our solution allows APs to save processing to enhance SE and evidence that the number of APs affects the SE differently in each network implementation.**

***Index Terms*—AP selection, cell-free massive MIMO networks, matched-decision, scalability, fine-tuning algorithms.**

I. Introduction

User-centric (UC) cell-free (CF) massive multiple-input multiple-output (MIMO) systems have emerged as a promising technology for future mobile communication networks. By employing many access points (APs) distributed over a coverage area, these systems provide higher macro-diversity against shadow fading and higher coverage probability than cell-based systems [1]–[5]. Besides, owing to the ability of each user equipment (UE) to connect to a subset of APs in its vicinity, the UC systems also demand fewer fronthaul/backhaul requirements than canonical CF systems, where all APs connect to all UEs [2], [3]. Even though the seminal papers embrace the UC approach, it can be impractical and unfair in some aspects, such as scalability, processing capabilities, and AP selection (also called AP clustering).

The distributed architecture employed by UC networks may demand high processing capabilities from the APs, since the APs need several advanced hardware components to process the signals of many UEs, such as a clock circuit and signal processor. Regarding scalability, the UC approach does not guarantee that the network resources, such as signal processing, signaling on fronthaul/backhaul, and total power are independent of the number of UEs [6]. Consequently, when the number of UEs is large, the network cannot provide a good service for any UE, making the UC approach unscalable [6], [7]. A possible strategy to partially solve these drawbacks is to restrict the maximum number of UEs that each AP can serve [2], [4]. Such an approach can also alleviate the processing capabilities demands from APs.

Regarding the AP selection process, even though the UC approach makes the UEs connect to partially overlapping subsets of APs, most previous works do not consider any cooperation among the APs in the AP cluster formation. They also do not limit the maximum number of UEs each AP can serve, making them unscalable [4]. Additionally, many of these AP selection methods are unfair since they do not prevent the worst UEs from being dropped. To the best of our knowledge, previous works also do not consider a matched-decision among the APs and the UEs, i.e., which connections among UEs and APs are more beneficial for both. They regularly assume that the UEs select a subset of APs to connect or that the APs select a subset of UEs to serve [2], [3]. However, relying on matched-decision AP selection is expected to significantly improve the system's performance. Moreover, the AP clusters generated by AP selection methods can comprise APs that contribute only marginally to the UE's performance, leading the APs to waste power with UEs that do not take advantage of the allocated power. Therefore, strategies that fine-tune the AP clusters of the UE while keeping the spectral efficiency (SE) under small degradation or improving energy efficiency (EE) are also indeed necessary.

A. Literature Review

The CF massive MIMO literature presents several AP selection schemes [8]–[12]. For instance, [8] associates the UE with

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).

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

Gilvan Borges is with Federal Institute of Pará - IFPA, Belém, PA, 66093-020 Brazil (e-mail: gilvan.borges@ifpa.edu.br).

Daniel Benevides da Costa is with the Technology Innovation Institute, 9639 Masdar City, Abu Dhabi, United Arab Emirates (email: danielbcosta@ieee.org).

André Mendes Cavalcante, Maria Marquezini and Igor Almeida are with Ericsson Research, Ericsson Telecomunicações S.A., Indaiatuba, SP, 13337-300 Brazil (e-mail: {andre.mendes.cavalcante;maria.marquezini;igor.almeida}@ericsson.com).

the AP, simultaneously presenting the largest channel gain and causing less interference. In contrast, [9] utilizes a graph neural network-based AP selection to reduce the number of reference signal received power (RSRP) measurements necessary to generate the AP clusters. Nonetheless, the complexity of [8] increases with the number of UEs, and the complexity of [9] can grow faster with the number of APs compared to other baseline solutions such as [4]. These are some strategies that the literature has proposed, but the most commonly analyzed methods are presented in [1]–[4]. In [1], the AP cluster of each UE is composed of all APs, which is the canonical version of CF. This method improves the SE of the worst UEs and increases the network's coverage probability compared to co-located systems. However, it can require exceptionally high computation costs and backhaul/fronthaul capabilities from the APs. Moreover, all APs have to divide their transmission powers among all UEs, leading the power allocation to be as small as zero in some cases. Therefore, the canonical form of CF is non-scalable (NS).

In [2], the APs serve the UEs with the largest large-scale fading in their vicinity. This scheme improves the SE of the UEs compared to canonical CF systems and avoids the depletion of allocated power by limiting the number of UEs each AP can serve. Nonetheless, it is an NS method since the APs need to have access to the channel statistics of all UEs in the network. Additionally, it does not prevent the worst UEs from being dropped.

In [3], the UE establishes a connection with the subset of APs that contribute most to the sum of its total channel gain, which can improve network energy efficiency. Nevertheless, it is also an NS scheme since it does not limit the number of UEs each AP can serve. Therefore, it does not guarantee scalability for the network resources. In [4], a scalable AP selection method that relates the pilot allocation to the cluster formation was proposed. The solution is based on the dynamic cooperation clustering (DCC) framework and prevents the worst UEs from being dropped. Moreover, the paper limits the number of UEs per AP and proves that only heuristic solutions that do not rely on the number of UEs are scalable. However, the literature in this research field is still in its infancy, and further investigations are indeed required to provide more valuable insights and advances in the area.

Therefore, those strategies generally present scalability issues, except [4], [9], and do not provide any mechanism to fine-tune the AP cluster of each UE. Moreover, they employ AP selection algorithms that aim to establish the best connections only for the UEs or the APs. Thus, this paper's main novelty is the proposal of a general AP selection framework that unifies the benefits of strategies whose AP-UE associations intend to provide the most suitable connections to the UEs and APs. Fine-tuning AP selection schemes are also another novelty of this paper.

B. Contributions

This paper proposes a general and scalable AP selection framework that exploits a matched-decision among UEs and APs. The method is divided into two stages, where the UEs first connect to an intermediate subset of APs and then to a final cluster of APs. The first stage allows the UEs and APs to establish the best connection for both in terms of large-scale fading. The second one enables the UEs to expand their AP clusters aiming to improve SE. Besides, modifying some parameters allows the proposed algorithm to behave like previous AP selection methods. However, it improves these schemes by affording scalability and reducing their time complexity.

Secondly, we propose novel methods for fine-tuning the AP clusters, cleverly dropping UE-AP connections that do not significantly contribute to the system performance. These are general strategies that work for any AP selection scheme. To the best of the authors' knowledge, this is the first paper that proposes general strategies for fine-tuning the AP clusters in scalable UC systems. Our analyses consider a wide range of UEs, APs, and antennas per AP. We assume that each AP can serve a limited number of UEs, where this number varies according to the scenario. We evaluate the system performance under different degrees of cooperation among the APs, i.e., centralized and distributed network implementations [13]. We compare our scheme with the canonical CF system and three baseline solutions, one of which is also a scalable approach. We also analyzed the system performance for perfect and imperfect knowledge of channel statistics. The contributions of this paper can be summarized as:

- We propose a general AP selection framework that considers a matched decision between the most advantageous connections for UEs and APs. The results indicate that the proposed method can improve the downlink (DL) SE of the 95% likely UEs up to 163% compared to the baseline solutions. Besides, it allows UC systems to employ APs to serve a small number of UEs and provide SEs as high as those whose APs can process the signals of more UEs.
- We provide scalable versions for the AP selection schemes presented in [2], [3] with negligible performance losses and lower time complexity.
- Three novel methods are proposed for fine-tuning the AP clusters. The first fine-tuning scheme relies on power allocation, the second is based on SE, and the third is on EE. Results indicate that the SE can be kept under minor degradation, even when dropping the UEs' connections with some APs.

C. Paper Outline and Notations

The remainder of this paper is organized as follows. Section II describes the system model, where the channel estimation procedure, the mathematical representation of the DCC framework, the network implementations, power allocation, and the DL SE are presented. Section III introduces the AP selection framework and the proposed fine-tuning schemes. Section IV plots illustrative numerical examples along with insightful discussions to show the effectiveness of the proposed approach compared to previous baseline schemes. Finally, Section V concludes the paper.

Notations: 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 expectation operator is denoted as $\mathbb{E}\{\,.\,\}$ and the cardinality of the set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The $N \times N$ identity matrix is $\mathbf{I}_N$ and $\mathrm{diag}(\mathbf{A}_1, ..., \mathbf{A}_n)$ denotes a block-diagonal matrix with the square matrices $\mathbf{A}_1, ..., \mathbf{A}_n$ on the diagonal. The euclidean norm is denoted as $\|\,.\,\|$ and $\mathrm{tr}(\,.\,)$ denotes the trace of a matrix. The complex Gaussian random variables (RVs) are denoted as $\mathcal{N}_{\mathbb{C}}\left(\mu, \sigma^2\right)$, where μ is the mean and σ^2 is the variance.

II. System Model

We consider a CF massive MIMO network composed of K single-antenna UEs and L APs distributed over the coverage area. Each AP is equipped with N antennas and the total number of antennas in the network is $M = NL$. The APs connect to the central processing units (CPUs) through fronthaul links, while the CPUs communicate with each other through backhaul ones. We assume that the network infrastructure (e.g., fronthaul, backhaul, and core network) is error-free and able to support the data traffic [1], [3], [14]. Moreover, the system operates on time-division-duplexing (TDD) protocol and it is assumed that the uplink (UL) and DL channels are reciprocal. Consequently, it is needed only to estimate the channel in the UL direction. Each coherence block comprises τ_c samples, where τ_c is divided into τ_p and τ_d samples for pilot signaling and DL data transmissions, respectively. The channel vector $\mathbf{h}_{kl} \in \mathbb{C}^{N\times 1}$ between the k-th UE and the l-th AP undergoes an independent correlated Rayleigh fading realization, which combines small-scale fading and statistical correlation matrix[1] $\mathbf{R}_{kl} \in \mathbb{C}^{N\times N}$. The channel vector $\mathbf{h}_{kl}$ can be computed as

$$\mathbf{h}_{kl} = \sqrt{\mathbf{R}_{kl}}\,\mathbf{g}_{kl}, \tag{1}$$

where $\mathbf{g}_{kl} \in \mathbb{C}^{N\times 1}$ is composed of elements that are independent and identically distributed (i.i.d) complex Gaussian $\mathcal{N}_{\mathbb{C}}\left(0, 1\right)$ RVs. We assume that the channels of different APs are uncorrelated since the APs are distributed over the coverage area [4], [15]. Moreover, we consider that the matrices $\mathbf{R}_{kl}$ remain unchanged for many coherence blocks, whereas the small-scale fading changes in each coherence block [16]. Additionally, one can note that for $l = \{1, ..., L\}$, $\mathbf{h}_{kl}$ and $\mathbf{R}_{kl}$ are sub-matrices of the collective vector $\mathbf{h}_k = \left[\mathbf{h}_{k1}^{\mathrm{T}}, ..., \mathbf{h}_{kL}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$ and the diagonal block matrix $\mathbf{R}_k = \mathrm{diag}\left(\mathbf{R}_{k1}, ..., \mathbf{R}_{kL}\right) \in \mathbb{C}^{M\times M}$, respectively.

A. Uplink Training

During the training phase, the UEs send pilot signals of τ_p-length to the APs to estimate their channels [14], [17]. The pilot signals are assumed to be mutually orthogonal and independent of K to provide scalability for the pilot resources. A pilot t can be reused by some UEs if $K > \tau_p$. Thus, letting $\mathcal{S}_t \subset \{1, ..., K\}$ represent the subset of UEs that send the pilot t, the received signal at l-th AP can be given by

$$\mathbf{y}_{tl}^{\mathrm{pilot}} = \sum_{i\in\mathcal{S}_t} \sqrt{\tau_p p_i}\,\mathbf{h}_{il} + \mathbf{n}_{il}, \tag{2}$$

[1]The statistical correlation matrix describes the large-scale fading of the channel, which is a function of antenna gains, shadowing, spatial channel correlation, and path loss [14].

where τ_p is the processing gain, p_i is the transmitted power of the i-th UE, and $\mathbf{n}_{il} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma^2\mathbf{I}_N\right)$ is the thermal noise. Assuming that the correlation matrix $\mathbf{R}_{kl}$ is perfectly known and utilizing the minimum mean-squared-error (MMSE) estimator, the estimated channel $\widehat{\mathbf{h}}_{kl}$ is given by

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p p_k}\,\mathbf{R}_{kl}\boldsymbol{\Psi}_{tl}^{-1}\mathbf{y}_{tl}^{\mathrm{pilot}}, \tag{3}$$

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

$$\boldsymbol{\Psi}_{tl} = \mathbb{E}\{\mathbf{y}_{tl}^{\mathrm{pilot}}(\mathbf{y}_{tl}^{\mathrm{pilot}})^{\mathrm{H}}\} = \sum_{i\in S_t} \tau_p p_i \mathbf{R}_{il} + \sigma^2\mathbf{I}_N. \tag{4}$$

The estimated channel $\widehat{\mathbf{h}}_{kl}$ relies on the perfect knowledge of $\mathbf{R}_{kl}$ and $\boldsymbol{\Psi}_{tl}$ in (3). However, channel statistics change in practice due to UE mobility or scheduling. Hence, each AP needs to estimate these matrices, leading $\widehat{\mathbf{h}}_{kl}$ to

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\tau_p p_k}\,\widehat{\mathbf{R}}_{kl}\widehat{\boldsymbol{\Psi}}_{tl}^{-1}\mathbf{y}_{tl}^{\mathrm{pilot}}, \tag{5}$$

where $\widehat{\mathbf{R}}_{kl}$ and $\widehat{\boldsymbol{\Psi}}_{tl}$ are the imperfect versions of $\mathbf{R}_{kl}$ and $\boldsymbol{\Psi}_{tl}$. We rely on [18] to calculate these terms, which proposes a two-stage approach. It considers that the l-th AP observes many channel realizations in different coherence blocks and uses the received pilots to estimate

$$\widehat{\boldsymbol{\Psi}}_{tl}^{(\mathrm{sample})} = \frac{1}{N_\Psi}\sum_{n=1}^{N_\Psi} \mathbf{y}_{tl}^{\mathrm{pilot}}[n]\left(\mathbf{y}_{tl}^{\mathrm{pilot}}[n]\right)^{\mathrm{H}}, \tag{6}$$

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

One can note that no extra-pilots are needed to calculate (6), since it is obtained from the pilots used to perform channel estimation. However, they are needed to compute $\widehat{\mathbf{R}}_{kl}$. To adjust this strategy to a CF scenario, we consider that τ_p pilots are utilized to estimate $\widehat{\mathbf{R}}_{kl}$ over N_R observations, requiring a total of $N_R\tau_p K$ extra pilots[2]. $\widehat{\mathbf{R}}_{kl}$ is computed as $\widehat{\mathbf{R}}_{kl} = \mu\widehat{\mathbf{R}}_{kl}^{(\mathrm{sample})} + (1-\mu)\widehat{\mathbf{R}}_{kl}^{(\mathrm{diagonal})}$, where μ is a regularization factor and $\widehat{\mathbf{R}}_{kl}^{(\mathrm{sample})} = \widehat{\boldsymbol{\Psi}}_{tl}^{(\mathrm{sample})} - \widehat{\boldsymbol{\Psi}}_{tl,-k}^{(\mathrm{sample})}$. The term $\widehat{\boldsymbol{\Psi}}_{tl,-k}^{(\mathrm{sample})}$ denotes a stage where only the interfering UEs sharing the same pilot as the k-th UE send the pilot. The parameters μ and η can be optimized according to the evaluated scenario when the APs are equipped with more than one antenna ($N > 1$), especially for $N \gg 1$. For simplicity, this paper evaluates the impacts of $\widehat{\mathbf{R}}_{kl}$ and $\widehat{\boldsymbol{\Psi}}_{tl}$ only for single antenna APs ($N = 1$). Hence, $\mu = 1$ and $\eta = 1$.

[2]The scalability of [18] is questionable as its complexity increases with K. However, this paper only investigates the influence of the imperfect knowledge of $\mathbf{R}_{kl}$ in the system performance. Therefore, a more profound discussion involving the best method of acquiring $\widehat{\mathbf{R}}_{kl}$ is out of the scope of this paper. For a deeper investigation, one can read the following references [19]–[21].

B. Matrix Representation of Dynamic Cooperation Clustering and Downlink Data Transmissions

In the DCC framework, the APs cooperate to create non-disjoint AP clusters that adapt to the network's time-varying conditions, such as channel properties and UE positions [22], [23]. To exploit the DCC concept and generate AP clusters that are dynamic and scalable, we proceed as follows. First, let $\mathcal{M}_k \subset \{1, ..., L\}$ denote the indexes of APs that serve the k-th UE. Second, let $\mathbf{c}_k = [\mathrm{c}_{1l}, ..., \mathrm{c}_{kL}] \in \mathbb{N}^{1\times L}$ be the vector which designates the APs that transmit a signal to the k-th UE. That is, if the l-th AP transmits a signal to the k-th UE, $\mathrm{c}_{kl} = 1$, otherwise $\mathrm{c}_{kl} = 0$, which means that

$$\mathrm{c}_{kl} = \begin{cases} 1 & \text{if } l \in \mathcal{M}_k \\ 0 & \text{if } l \notin \mathcal{M}_k \end{cases}. \tag{7}$$

Moreover, the matrix $\mathbf{D}_{kl} \in \mathbb{N}^{N\times N}$ is used to describe which antennas of the l-th AP establish a connection to the k-th UE. It is assumed that all N antennas of the l-th AP transmit a signal to the k-th UE. This is similar to write [4], [6]

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

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

The cardinality of $\mathcal{U}_l$ is regularly lower than K. However, if all UEs group in the vicinity of specific APs, $|\mathcal{U}_l|$ can be no longer $|\mathcal{U}_l| < K$. Conversely, $|\mathcal{U}_l|$ may be equal to K, which is unscalable. However, if all UEs group in the vicinity of specific APs, $|\mathcal{U}_l|$ may be equal to K, which is unscalable. In order to solve this drawback, this paper assumes that $|\mathcal{U}_l| \leq U_{max}$, where U_{max} is a constant that remains unchanged even if $K \to \infty$. That is, even if the number of UEs is extremely high, the APs will serve at most U_{max} UEs in $\mathcal{U}_l$. The constant U_{max} can also be seen as a type of processing capacity limitation regarding the maximum number of UEs that each AP can process signals. Section III will present with further details the proposed methodology to acquire $\mathbf{c}_k$ and $\mathbf{D}_k$. Let $\mathbf{x}_l = \sum_{k=1}^{K} \mathbf{D}_{kl}\mathbf{w}_{kl}s_k$ denote the data signal sent by the l-th AP. The DL received signal is given by [6]

$$y_k^{dl} = \underbrace{\mathbf{h}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{w}_k s_k}_{\text{Desired signal}} + \sum_{i=1, i\neq k}^{K} \underbrace{\mathbf{h}_k^{\mathrm{H}}\mathbf{D}_i\mathbf{w}_i s_i}_{\text{Interfering signals}} + \underbrace{n_k}_{\text{Noise}}, \tag{9}$$

where $s_i \in \mathbb{C}$ is the signal transmitted for the i-th UE, which satisfies $\mathbb{E}\left\{\|s_i\|^2\right\} = 1$; $n_k \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma_{dl}^2\right)$ is the receiver noise, $\mathbf{w}_{il} \in \mathbb{C}^{N\times 1}$ is the precoding vector, and $\mathbf{w}_i = \left[\mathbf{w}_{i1}^{\mathrm{T}}, ..., \mathbf{w}_{iL}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$ is the collective precoding vector.

C. Network Implementations

We consider two network implementations, a centralized and a distributed [13]. In the centralized one, the CPUs take the responsibility to estimate the channels, apply precoding, and process the DL signals. This implementation provides better interference cancellation, and the CPUs have access to channel statistics. In the distributed implementation the processing tasks are performed locally in the APs using local channel state information (CSI). The only task of the CPUs is the encoding of the DL data signals. This implementation may require less signaling on the fronthaul/backhaul links[3].

To compute the precoding vectors in a scalable way, the UL/DL duality is considered. In the centralized implementation, the precoding vector is selected as $\mathbf{w}_k$ in (10), where $\overline{\mathbf{w}}_k \in \mathbb{C}^{M\times 1}$ determines the direction of the precoding vector, ϱ_k is the power allocated to the k-th UE, and $\mathbb{E}\left\{\|\mathbf{w}_k\|^2\right\} = \varrho_k$. For the distributed implementation, $\mathbf{w}_{kl}$ is selected on the basis of its local channel estimate as also shown in (10), where ϱ_{kl} is the power that the l-th AP allocates for the k-th UE. We adopt the maximum-ratio (MR) and local partial minimum mean-squared-error (LP-MMSE) precoding for the distributed implementation, while we employ the partial regularized zero-forcing (P-RZF) and partial minimum mean-squared-error (P-MMSE) for the centralized one, due to the scalability features of these precodings, which yields [4]

$$\mathbf{w}_k = \sqrt{\varrho_k}\frac{\overline{\mathbf{w}}_k}{\sqrt{\mathbb{E}\left\{\|\overline{\mathbf{w}}_k\|^2\right\}}}, \mathbf{w}_{kl} = \sqrt{\varrho_{kl}}\frac{\overline{\mathbf{w}}_{kl}}{\sqrt{\mathbb{E}\left\{\|\overline{\mathbf{w}}_{kl}\|^2\right\}}}. \tag{10}$$

The scalability aspects in LP-MMSE, P-MMSE, and P-RZF are indicated by the term partial (acronym P). It suggests that they only use the channel estimates and statistics of a subset of UEs to mitigate interference. Moreover, it is worth noting that all precoding vectors are calculated under imperfect CSI. For instance, $\overline{\mathbf{w}}_k = \widehat{\mathbf{h}}_k$ and $\overline{\mathbf{w}}_{kl} = \widehat{\mathbf{h}}_{kl}$ for the MR precoding. Two heuristic methods for power allocation are considered to address scalability aspects in both network implementations. We employ fractional power allocation[4] for the centralized one, since it is a method that keeps the ability to mitigate interference and performs better than equal power allocation for the worst UEs. Thus, ϱ_k is a function of β_{kl}, $\upsilon \in [-1, 1]$, $\kappa \in [0, 1]$, and ϱ_l, where ϱ_l is the total transmission power of each AP and β_{kl} denotes the large-scale fading of the k-th UE to the l-th AP. Moreover, υ and κ are project parameters [6]. For the distributed one, we use the method that divides the power resources proportionally to the large-scale fading gains of each UE [26]. Thus, ϱ_{kl} is given by

$$\varrho_{kl} = \begin{cases} \varrho_l \frac{\sqrt{\beta_{kl}}}{\sum_{i\in\mathcal{U}_l}\sqrt{\beta_{il}}} & \text{if } k \in \mathcal{U}_l \\ 0 & \text{otherwise} \end{cases}, \tag{11}$$

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

[3]This is not always true. If $\tau_c/(\tau_c - \tau_p) \approx 1$ and $K \gg N$, the distributed implementation may require much more signaling [13].

[4]There is a wide range of power allocation methods in the literature [24]–[26]. However, analyses involving the best power allocation algorithm are out of the scope of this paper.

D. Spectral and Energy Efficiencies

In order to calculate the SE of DL channels, we rely on the received signal presented in (9) so that an achievable DL SE for the k-th UE can be expressed as [6]

$$\mathrm{SE}_k = P_f \log_2 (1 + \mathrm{SINR}_k), \tag{12}$$

where P_f is the pre-log factor, which is a fraction of samples per coherence block that is used to transmit the DL data. For perfect knowledge of correlation matrices, $P_f = \tau_d/\tau_c$ and $P_f = 1 - (\tau_p/\tau_c) - \alpha$ for imperfect knowledge, where $\alpha = N_R \tau_p K/\tau_s \tau_c$. The term SINR_k denotes the signal-to-interference-plus-noise ratio (SINR) in the DL direction. From (9), the SINR_k can be computed as

$$\mathrm{SINR}_k = \frac{|\mathrm{DS}_k|^2}{\mathrm{IS}_k - |\mathrm{DS}_k|^2 + \sigma_{dl}^2}. \tag{13}$$

where $\mathrm{DS}_k = \mathbb{E}\left\{\mathbf{h}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{w}_k\right\}$ denotes the desired signal and $\mathrm{IS}_k = \sum_{i=1}^{K} \mathbb{E}\left\{\left|\mathbf{h}_k^{\mathrm{H}} \mathbf{D}_i \mathbf{w}_i\right|^2\right\}$ stands for the interference. Eq. (12) is also known as hardening bound, which is commonly used in massive MIMO theory and is valid for any choice of precoding vectors [13], [25]. It can be seen as a capacity lower bound and, unfortunately, it does not have a closed-form expression when using P-MMSE and LP-MMSE, but can be computed through Monte-Carlo simulations if $\mathbf{w}_i$ is selected as in (10). Besides, all expectations presented in (13) are related to the channel realizations [6].

The total EE in bit/Joule is defined as the ratio between the total data rate $R_t = B_s \sum_{k=1}^{K} \mathrm{SE}_k$, and the total power consumption in Watts [27]. It can be defined as [3]

$$\mathrm{EE}_t = \frac{R_t}{\sum_{l=1}^{L} \left\{ \frac{1}{\nu_l} \mathbb{E}\left\{\|\mathbf{x}_l\|^2\right\} + N P_{\mathrm{tc},l} + P_{\mathrm{fh},l} \right\}}, \tag{14}$$

where $0 < \nu_l \leq 1$ is the efficiency of the power amplifier, and $P_{\mathrm{tc},l}$ is the power that each antenna of the AP needs to run internal components, such as converters and filters. Besides, $P_{\mathrm{fh},l}$ is the power consumption in the fronthaul link connecting a CPU and an AP, given by $P_{\mathrm{fh},l} = P_{0,l} + P_{\mathrm{ft},l} B_s \sum_{k \in \mathcal{U}_l} \mathrm{SE}_k$, where $P_{0,l}$ is a fixed power consumption of each link and $P_{\mathrm{ft},l}$ is the traffic-dependent power in Watt per bit/s.

III. Proposed AP Selection Framework and Fine-Tuning Schemes

We propose a novel AP selection framework that exploits a competitive mechanism and considers a matched decision among UEs and APs. The scheme is divided into two stages, and the flowchart exhibited in Fig. 1 provides an overview of the method's operation. The matched-decision process occurs in the first stage (named intermediate AP cluster). The UEs connect to an intermediate subset of APs, aiming to make the UEs and APs establish the best connection for both in terms of large-scale fading. In the second stage (called final AP cluster), the UEs try to connect to more APs and expand their AP clusters, intending to improve the SE. For clarification, Figs. 2 (a) and 2 (b) illustrate the intermediate and final AP clusters, respectively. The first stage can enable better use of the power resources since the matched-decision allows the APs to serve the best UEs in their vicinity. The second one gives the worst UEs a chance to increase their DL SE.

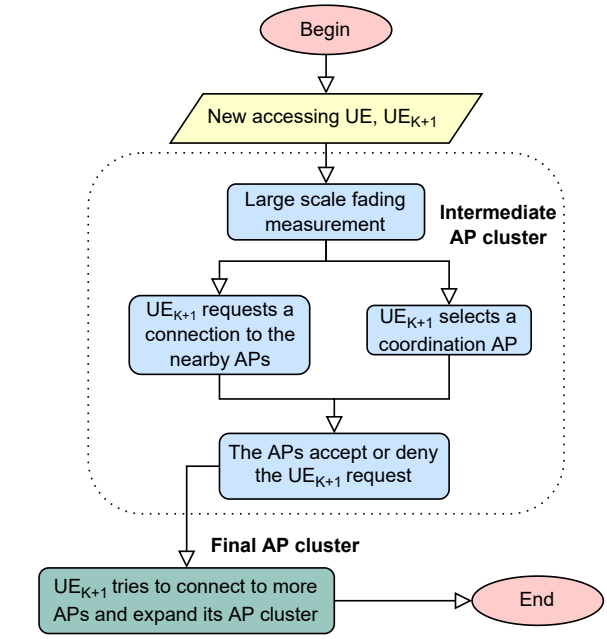


Fig. 1: Flowchart of the proposed matched-decision AP selection method.

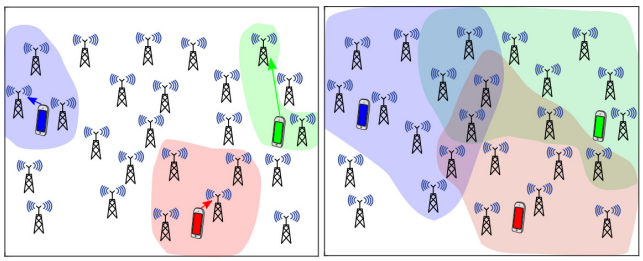


Fig. 2: Illustration of the AP cluster formation. (a) Intermediate AP clusters, with arrows indicating the coordination APs. (b) Final AP clusters after the second stage.

We assume a specific limitation in the processing capacity of the APs, that is, each AP can serve only a limited number of UEs, named U_{max}. However, unlike previous works, we do not consider that U_{max} is always equal to τ_p [4], [6], [7]. Instead, we assume that U_{max} depends on the AP processing capabilities, such that $1 \leq U_{max} \leq \tau_p$. Additionally, the AP selection is generated on a per-UE basis to achieve scalability, such that the AP clustering occurs only among the UEs and the APs. The CPUs do not participate in the AP cluster formation. For simplicity, we denote a new accessing UE by the notation UE_{K+1} instead of $(K+1)$-th UE.

A. Intermediate AP Cluster

Fig. 2 (a) illustrates the intermediate AP clusters created for each UE in the network. Its generation is detailed as follows: when a new UE_{K+1} enters the network, it measures the large-scale fading $\beta_{(K+1)l}$ of the APs in its vicinity. Then, the

UE requests a connection to a subset of APs according to a decision criterion that follows the requirements of the system's design. This paper considers that the UE_{K+1} will request a connection only to the APs whose channel gains satisfy $\beta_{(K+1)l} > \gamma$. The target is to evaluate the system performance under a controlled parameter (γ), where γ refers to a threshold gain. The subset of APs selected by the UE_{K+1} is denoted by $\mathbf{e}_{(K+1)} = \left[\mathrm{e}_{1l}, ..., \mathrm{e}_{(K+1)L}\right] \in \mathbb{N}^{1\times L}$ and is defined as

$$\mathrm{e}_{(K+1)l} = \begin{cases} 1 & \text{if } \beta_{(K+1)l} > \gamma \\ 0 & \text{otherwise} \end{cases}. \tag{15}$$

However, although UE_{K+1} desires to connect to the APs that meet the decision criterion, the connection will only be carried out whether these APs individually accept the request of the UE_{K+1}. That is, there must be a matched-decision between the UE_{K+1} and the APs. In other words, the connection must be advantageous for both UE_{K+1} and APs. The APs can employ several decision criteria, such as the least pilot contamination [4], effective channel gain [8], among others. Nonetheless, to use similar criteria across UEs and APs, we assume that the decisions rely on channel gain $\beta_{(K+1)l}$ in the APs. The decisions are denoted by $\mathbf{f}_{(K+1)} = \left[\mathrm{f}_{1l}, ..., \mathrm{f}_{(K+1)L}\right] \in \mathbb{N}^{1\times L}$ and are summarized as

$$\mathrm{f}_{(K+1)l} = \begin{cases} 1 & \text{if } \beta_{(K+1)l} > \beta_{il}^{min} \\ 0 & \text{otherwise} \end{cases}, \tag{16}$$

where $i \neq (K+1)$ denotes the UE with the smallest channel gain (β_{il}^{min}) that the l-th AP serves in $\mathcal{U}_l$. In case of $\beta_{(K+1)l} > \beta_{il}^{min}$, the AP accepts the UE_{K+1} and keeps the connection of the i-th UE only if there are available connections (i.e., $|\mathcal{U}_l| < U_{max}$). The intermediate AP cluster is given by

$$\mathbf{c}_{(K+1)}^{int} = \mathbf{e}_{(K+1)} \wedge \mathbf{f}_{(K+1)}, \tag{17}$$

where $\wedge$ is the logical operation AND. From (17), one can note that the UE_{K+1} may not connect with any AP if it is rejected by all APs in (16). To circumvent this issue, the UE_{K+1} also claims a coordination AP. This AP ensures that the network will no longer drop the UE_{K+1} since it serves the UE regardless of its channel condition [4]. We consider that each UE has a coordination AP and that its choice is independent of the threshold γ. Hence, the UE connects to at least one AP. Consequently, the cardinality of the subset of APs that serve the UE_{K+1} becomes

$$|\mathcal{M}_{K+1}| = \sum_{l \in \mathcal{M}_{K+1}} \mathrm{c}_{(K+1)l}^{int} \geq 1. \tag{18}$$

In Fig. 2 (a), the arrows indicate the coordination AP of each UE. This paper assumes that the procedure for choosing a coordination AP is the same for all UEs. Thus, the procedure will be explained only for the UE_{K+1}. To choose a coordination AP, the UE_{K+1} solicits a connection to the available APs. Then, the available APs respond, and the UE_{K+1} selects the one with the strongest channel gain $\beta_{(K+1)l}$ to be its coordination AP. Let $\mathcal{A}_l \subset \mathcal{U}_l$ denote the subset of UEs that the l-th AP is coordinating. The available APs are the ones presenting $|\mathcal{A}_l| < U_{max}$, $\forall l \in \{1, \ldots, L\}$. That is, the APs that are not using their total processing capacity only for coordinating UEs. Therefore, the coordination AP of the UE_{K+1} can be defined by

$$\begin{aligned} l = \arg\max_l \beta_{(K+1)l} \\ \text{s.t}\, |\mathcal{A}_l| < U_{max} \end{aligned}, \tag{19}$$

which means that although several APs (that present $|\mathcal{A}_l| < U_{max}$) reply to the request of the UE_{K+1}, it selects only the AP with the highest $\beta_{(K+1)l}$ to be its coordination AP. It is worth noting that in (19), $|\mathcal{A}_l| + |\mathcal{B}_l| \leq U_{max}$, where $\mathcal{B}_l \subset \mathcal{U}_l$ denotes the subset of UEs that the l-th AP serves, but do not coordinates (i.e., UEs that the AP can drop). After that, the coordination AP also informs other APs that a new accessing UE is in the coverage area. Due to the addition of the coordination AP, (15) becomes

$$\mathrm{e}_{(K+1)l} = \begin{cases} 1 & \text{if } \beta_{(K+1)l} > \gamma \\ 1 & \text{if } \mathrm{UE}_{K+1} \in \mathcal{A}_l \\ 0 & \text{otherwise} \end{cases}, \tag{20}$$

where $\mathcal{A}_l \subset \mathcal{U}_l$. Additionally, (16) also changes to

$$\mathrm{f}_{(K+1)l} = \begin{cases} 1 & \text{if } \beta_{(K+1)l} > \beta_{il}^{\min} \\ 1 & \text{if } \mathrm{UE}_{K+1} \in \mathcal{A}_l \\ 0 & \text{otherwise} \end{cases}, \tag{21}$$

where $i \neq (K+1)$ and $i \in \mathcal{B}_l$. One can note that the network will no longer drop the UE_{K+1} if we apply (20) and (21) in (17) instead of (15) and (16), since the choice of the coordination AP is independent of γ. It is worth noting that the subset $\mathcal{B}_l$ does not affect the coordination AP assignment in (19). For instance, if $|\mathcal{U}_l| = U_{max}$ and $|\mathcal{B}_l| \geq 1$, the l-th AP could drop the UE with the weakest channel gain belonging to subset $\mathcal{B}_l$ (UEs not coordinated) in order to coordinate the UE_{K+1} in subset $\mathcal{A}_l$. Such an approach guarantees that $|\mathcal{U}_l| \leq U_{max}$ since $|\mathcal{A}_l| + |\mathcal{B}_l| \leq U_{max}$. Algorithm 1 summarizes the intermediate AP selection process.

Algorithm 1: Intermediate AP cluster

Input: γ, U_{max}, $i = 1, ..., K$, $l = 1, ..., L$, $k = K+1$;
The UE connects to a coordination AP by solving (19);
Update $\mathcal{A}_l$ according to the solution of (19);
for $l = 1$ **to** L **do**
  $\mathrm{e}_{kl} = 0$; $\mathrm{f}_{kl} = 0$; $\mathrm{c}_{kl} = 0$;
  // k-th UE requests connections to nearby APs:
  if $k \in \mathcal{A}_l$ **or** $\beta_{kl} > \gamma$ **then**
    $\mathrm{e}_{kl} = 1$;
  end
  // APs accept or reject the UE request:
  if $k \in \mathcal{A}_l$ **or** $\beta_{kl} > \beta_{il}^{min}$ **then**
    $\mathrm{f}_{kl} = 1$; $\mathrm{f}_{il} = 1$; // where $i \in \mathcal{B}_l$
    if $|\mathcal{U}_l| = \mathrm{U}_{max}$ **then**
      $\mathrm{f}_{il} = 0$;
    end
  end
  $\mathrm{c}_{kl}^{int} = (\mathrm{e}_{kl} \wedge \mathrm{f}_{kl})$ // Matched-decision
end
Output: $\mathbf{c}_k^{int} = \left[\mathrm{c}_{k1}^{int}, \ldots, \mathrm{c}_{kL}^{int}\right]$.

B. Final AP Cluster Formation

Fig. 2 (b) depicts the final AP clusters of each UE. Its generation is detailed as follows: after the formation of vectors $\mathbf{e}_{(K+1)}$ and $\mathbf{f}_{(K+1)}$, the UE_{K+1} may have been rejected by some APs. However, the UE has a new chance to link to these APs and then expand its AP cluster in this second step. To do this, a j-th AP that initially rejected the UE_{K+1}, verify if there are still connections available for a new UE. That is, it checks if $|\mathcal{U}_j| < U_{max}$. Then, the AP accepts the request of the UE_{K+1} if this condition is satisfied. Computationally, this can be represented by the vector $\mathbf{z}_{(K+1)j} = \left[\mathrm{z}_{1l}, ..., \mathrm{z}_{(K+1)L}\right] \in \mathbb{N}^{1\times L}$, which is given by

$$\mathrm{z}_{(K+1)j} = \begin{cases} 1 & \text{if } |\mathcal{U}_j| < U_{max} \\ 0 & \text{otherwise} \end{cases}, \tag{22}$$

where $j \neq l$. Nevertheless, the UE_{K+1} may be dropped again in (21) if a new UE with a better channel condition enters the network. In the end, the final AP cluster is computed as

$$\mathbf{c}_{(K+1)} = \mathbf{c}^{int}_{(K+1)} \vee \mathbf{z}_{(K+1)}, \tag{23}$$

where the operator $\vee$ denotes the logical operation OR. From (23), one can compute the number of APs serving the UE_{K+1} by calculating $\left|\mathcal{M}_{(K+1)}\right| = \sum_{l\in\mathcal{M}_{(K+1)}} \mathbf{c}_{(K+1)}$. One can also compute $\mathbf{D}_{(K+1)}$ through $\mathbf{c}_{(K+1)}$ by assuming that $\mathbf{D}_{(K+1)l} = \mathbf{I}_N$, when $\mathrm{c}_{(K+1)l} = 1$. Otherwise $\mathbf{D}_{(K+1)l} = \mathbf{0}_N$, for $l = \{1, ..., L\}$. For clarification, Algorithm 2 summarizes the final AP clustering.

Algorithm 2: Final AP cluster

Input: $\mathbf{c}^{int}_k$, U_{max}, $l = 1, ..., L$, $k = K + 1$;
for $l = 1$ **to** L **do**
  // Connect the k-th UE with available APs:
  if $|\mathcal{U}_l| < U_{max}$ **then**
    $\mathrm{z}_{kl} = 1$;
  end
  $\mathrm{c}_{kl} = \mathrm{c}^{int}_{kl} \vee \mathrm{z}_{kl}$;
end
Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

The AP selection proposed in this paper is composed of the Algorithms 1 and 2. In the first step of the Algorithm 1, the time complexity for each new UE to select its coordinating AP by solving (19) is $\mathcal{O}(L)$. Then, the complexity for each new UE that requests connections to nearby APs is $\mathcal{O}(|\mathcal{A}_l|)$, while for each AP to accept or reject the request is $\mathcal{O}(|\mathcal{B}_l|)$. This comes from the fact that only the UEs belonging to subsets $\mathcal{A}_l$ and $\mathcal{B}_l$ are evaluated. The complexity for computing the intermediate AP cluster is $\mathcal{O}(L(|\mathcal{A}_l|+|\mathcal{B}_l|))$, since these steps are repeated for each AP. Then, noticing that $|\mathcal{A}_l|+|\mathcal{B}_l| \leq |\mathcal{U}_l|$, the complexity of Algorithm 1 simplifies to $\mathcal{O}(L|\mathcal{U}_l|)$. The complexity to compute the final AP cluster in Algorithm 2 is also $\mathcal{O}(L|\mathcal{U}_l|)$, since only the UEs belonging to subset $|\mathcal{U}_l|$ are evaluated for each AP. Therefore, the overall time complexity of the proposed AP selection is $\mathcal{O}(L|\mathcal{U}_l|)$, which can be further simplified to $\mathcal{O}(LU_{max})$ by noticing that $|\mathcal{U}_l| \leq U_{max}$. Therefore, one can note the scalability aspect of the proposed strategy since its complexity does not increase with K.

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

C. Comparison with other AP Selection Methods

The proposed solution aims to generate AP clusters composed of the more convenient connections for UEs and APs. Consequently, it inherits several characteristics of classical AP selection schemes (decisions taken only in the UEs or in the APs) and can degenerate into these by adjusting the vectors $\mathbf{e}_{(K+1)}$, $\mathbf{f}_{(K+1)}$, and $\mathbf{z}_{(K+1)}$. Before showing it, let us briefly describe some solutions we use for comparisons.

1) Canonical CF [1]: it is a NS scheme in which the AP cluster of each UE is composed of all APs. This method improves the SE of the worst UEs and increases the network's coverage probability compared to co-located systems.

2) User-centric clustering (UCC) [2]: the APs serve the U_{max} UEs with the greatest large-scale fading in their vicinity. It is a NS method that does not prevent the worst UEs from being dropped. The time complexity of this method is $\mathcal{O}(LK \log K)$ for each UE as whenever a new UE enters the network, all APs have to perform a sorting operation to select U_{max} UEs to serve.

3) Largest-large-scale-fading-based (LSFB) [3]: the UE establishes a connection with the subset of APs that contribute most to the sum of its total channel gain, in percentage $\delta\%$. This strategy has complexity $\mathcal{O}(L \log L)$ for each UE, as each new UE sorts the channel gains of L APs during the AP selection process. One can note that this method could be scalable, as its complexity does not increase with K. Nevertheless, it is also a NS scheme since it does not limit the number of UEs that each AP can serve.

4) Scalable CF [4]: the UE connects to master and non-masters APs and the AP selection is intrinsically related to the pilot assignment. Among the reference schemes analyzed, this is the only one that is scalable and also prevents the worst UEs from being dropped. However, providing a mechanism that aims further to improve the SE of the worst UEs is not one of the goals of this method. The time complexity is $\mathcal{O}(L\tau_p)$ for each UE, which is due to the fact that after the pilot assignment process, each AP chooses to serve up to one UE per pilot. The complexity of the pilot assignment method for all UEs K is $\mathcal{O}((K-\tau_p)\tau_p)$. In this method, the UE chooses a master AP, and the APs performs the AP cluster expansion.

The matched-decision scheme behaves like a UCC strategy if we solve (19) and set $\mathrm{e}_{(K+1)l} = 1$ for all APs, which is similar to considering a small γ in (20). Hence, the UEs' choices do not impact (17), and the APs' decisions dominate the AP cluster formation. Consequently, the APs tend to

select the UEs presenting the best channel gain in their vicinity in (21), leading to a UCC scheme. However, this UCC implementation achieves scalability, guarantees connection for all UEs, and its time complexity is independent of the number of UEs. Hereafter, we name it matched-decision (MD) UCC.

The proposed scheme can also behave like a scalable version of the LSFB algorithm. Hereafter, we name it MD LSFB. To this end, one should solve (19) and consider $\mathrm{e}_{(K+1)l} = 1$ for the subset of APs that contribute most to the total sum channel gain $\delta\%$, instead of relying on γ in (20). One can also achieve a similar implementation of the scalable CF scheme by solving (19) and considering $\mathrm{e}_{(K+1)l} = 1$ and $\mathrm{z}_{(K+1)l} = 0$ for all APs. Then, (21) should be modified to make the APs serve only the UEs, causing the least pilot contamination. Regarding the canonical CF, it is only a particular case of the UC approach when $\mathrm{c}_{kl} = 1$ and $\mathbf{D}_{kl} = \mathbf{I}_N$ for all APs and UEs [4].

D. Fine-Tuning AP Selection

The AP cluster of each UE created by AP selection schemes can comprise APs that contribute only marginally to the UE's performance. Therefore, we propose two strategies to reduce the number of APs connected to each UE while avoiding reducing the SE significantly, and a third one that aims to improve the EE. The first one is performed locally in each AP, without the CPUs participation. The l-th AP can drop uncoordinated UEs (i.e., UEs that are in $\mathcal{B}_l$) that receive only a small fraction of the total power in (11). The AP sorts $\{\varrho_{1l}, ..., \varrho_{kl}\}$ in descending order to identify the uncoordinated UEs that receive more power, leading to $\{\bar{\varrho}_{1l}, ..., \bar{\varrho}_{k'l}\}$. The indexes of the unsorted UEs are stored in the k'-th element of the subset $\bar{\mathcal{B}}_l$. Then, the AP carries out a cumulative sum, which can be expressed as

$$\varrho^{\mathrm{sum}}_{k'l} = \begin{cases} \frac{\bar{\varrho}_{k'l}}{\sum_{k\in\mathcal{U}_l}\varrho_{kl}} & \text{if } k' = 1 \\ \frac{\bar{\varrho}_{k'l}}{\sum_{k\in\mathcal{U}_l}\varrho_{kl}} + \varrho^{\mathrm{sum}}_{(k'-1)l} & \text{otherwise} \end{cases}. \tag{24}$$

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

The second fine-tuning scheme works in a centralized fashion, and it is based on SE. Specifically, the CPUs drop the connections of APs that contribute only marginally to the SE of the k-th UE. For fine-tuning the AP clusters, the CPUs consider that all APs apply MR precoding and assume perfect CSI to simplify the calculation of the desired signal (DS_k) and interference (IS_k) terms in (13), which are computed as

$$\begin{aligned} \mathrm{DS}_k &= \mathbb{E}\left\{\mathbf{h}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{w}_k\right\} = \sum_{l\in\mathcal{M}_k}\sqrt{\varrho_k \,\mathrm{tr}\left(\mathbf{R}_{kl}\right)} \\ \mathrm{IS}_k &= \sum_{i\in\mathcal{P}_k}\mathbb{E}\left\{\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{D}_i\mathbf{w}_i\right|^2\right\} = \sum_{i\in\mathcal{P}_k}\varrho_i\sum_{l\in\mathcal{M}_i}\frac{\mathrm{tr}\left(\mathbf{R}_{il}\mathbf{R}_{kl}\right)}{\mathrm{tr}\left(\mathbf{R}_{il}\right)}, \end{aligned} \tag{25}$$

where $\mathcal{P}_k$ denotes the subset of UEs that are partially served by the same APs as the k-th UE. $\mathcal{P}_k$ is adopted to make (25) scalable since by definition $\mathcal{P}_k = \{i : \mathbf{D}_k\mathbf{D}_i \neq \mathbf{0}_{LN\times LN}\}$ [6]. Next, the CPUs estimate SINR_k in (13) and calculate the SE of the k-th UE in (12).

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

$$\bar{\mathrm{q}}^{\mathrm{sum}}_{kl'} = \begin{cases} \bar{\mathrm{q}}_{kl'} & \text{if } l' = 1 \\ \bar{\mathrm{q}}_{kl'} + \bar{\mathrm{q}}^{\mathrm{sum}}_{k(l'-1)} & \text{otherwise} \end{cases}, \tag{26}$$

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

Algorithm 3: Fine-tuning based on power allocation

Input: $\Gamma\%$, $\mathcal{B}_l$, $k = 1, ..., |\mathcal{B}_l|$;
Sort ϱ_{kl} in descending order for the uncoordinated UEs
for $k' = 1$ **to** $\left|\bar{\mathcal{B}}_l\right|$ **do**
  Perform a cumulative sum in (24)
  Map k' to the unsorted value of k in the subset $\bar{\mathcal{B}}_l$
  if $\varrho^{sum}_{k'l} \leq \Gamma\%$; **then**
    $\mathrm{c}_{kl} = 1$ // Performed in the APs
  else
    $\mathrm{c}_{kl} = 0$
  end
end
Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

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

One can note that only channel statistics are needed to compute the proposed fine-tuning schemes, making them valid for many coherence blocks. In (25), the CPUs have to perform about $|\mathcal{P}_k|\,|\mathcal{M}_i|\,N^3$ complex multiplications per UE, and for scalability purposes, we assume that the CPUs can fine-tune the AP clusters of only LU_{max} UEs, corresponding to the number of connections in the network.

The fine-tuning scheme presented in Algorithm 4 also works if the CPUs consider the imperfect CSI to calculate DS_k and IS_k. In this case, the closed-form expressions of SE

Algorithm 4: Fine-tuning based on SE

Input: ε, $\mathcal{M}_k$, $l = 1, ..., |\mathcal{M}_k|$;
Compute SE_k in (12) using (25) and create $\mathbf{q}_k$
Sort the elements of $\mathbf{q}_k$ in descending order
for $l' = 1$ **to** $\left|\bar{\mathcal{M}}_k\right|$ **do**
  Perform the cumulative sum in (26)
  $\mathrm{SINR}_{kl'}^{\mathrm{sum}} = \left|\bar{\mathrm{q}}_{kl'}^{\mathrm{sum}}\right|^2 / (\mathrm{IS_k} - \left|\bar{\mathrm{q}}_{kl'}^{\mathrm{sum}}\right|^2 + \sigma_{\mathrm{dl}}^2)$
  Using $\mathrm{SINR}_{kl'}^{\mathrm{sum}}$, compute $\mathrm{SE}_{kl'}^{\mathrm{sum}}$ in (12)
  Map l' to the unsorted value of l in the subset $\bar{\mathcal{M}}_k$
  if $\mathrm{SE}_k - \mathrm{SE}_{kl'}^{\mathrm{sum}} \geq \varepsilon$; **then**
    $\mathrm{c}_{kl} = 1$ // Performed in the CPUs
  else
    $\mathrm{c}_{kl} = 0$
  end
end
Output: $\mathbf{c}_k = [\mathrm{c}_{k1}, \ldots, \mathrm{c}_{kL}]$.

derived for MR in [4] could be employed, but at the cost of $|\mathcal{P}_k|\,|\mathcal{M}_i|\,(8N^3 - N)$ complex multiplications per UE. Moreover, a pilot assignment strategy has to be performed before the fine-tuning process. It is essential to emphasize that the CPUs assume perfect CSI only to simplify the calculation of DS_k and IS_k. Recall that perfect CSI is not employed in the precoding vectors.

Algorithm 5: Fine-tuning based on EE

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

The third proposed strategy aims to improve EE. It fine-tunes the AP cluster based on the SE and allocated power ratio ($\mathrm{SE}_k/\varrho_{kl}$). This strategy aims to remove the connections with UEs that achieve high SE and consume a small percentage of the AP power resources. Although this approach may seem counter-intuitive, note that the APs allocate more power (ϱ_{kl}) to the UEs presenting the strongest channel gains in (11). Besides, SE_k increases with the desired signal, which is proportional to the UE channel gain in each serving AP, as (25) demonstrates. Thus, if a UE achieves a high SE, receiving only a tiny fraction of power from the AP compared to other UEs, it indicates that this AP is not so fundamental to the SE of this UE. In (14), one can also note that the power consumption on fronthaul links ($P_{\mathrm{fh},l}$) is proportional to the SE of the UEs the AP serves in $\mathcal{U}_l$. Therefore, some UEs do not benefit as much from some APs, but contribute to increasing $P_{\mathrm{fh},l}$.

To fine-tune the AP clusters, we proceed as follows: the power allocation is performed in (11), and the SE of each UE is computed in (12) using (13) and (25). In the following, the CPUs find the UE presenting the maximum ratio $\mathrm{SE}_k/\varrho_{kl}$ in each AP, such that $k_{(l,max)} = \arg\max_k (\mathrm{SE}_k/\varrho_{kl})$ and $\mathrm{r}_{k_{(l,max)}} = \mathrm{SE}_{k_{(l,max)}}/\varrho_{k_{(l,max)}l}$. For scalability purposes, we adopt a heuristic solution for making the APs drop uncoordinated UEs (i.e., UEs that are in $\mathcal{B}_l$). First, it is assumed that $\mathrm{c}_{k_{(l,max)}l} = 0$. Then, we consider that $\mathrm{c}_{kl} = 1$ if $\mathrm{SE}_k/\varrho_{kl} < \zeta \mathrm{r}_{k_{(l,max)}}$, and $\mathrm{c}_{kl} = 0$ otherwise, where ζ is a project parameter. Each CPU performs these tasks for the APs linked to it by fronthaul. The elements of $\mathcal{B}_l$ are indexed by k'' and Algorithm 5 summarizes the entire process.

IV. Numerical Results

We consider a CF network consisting of L APs, each equipped with N antennas. Each AP can serve up to U_{max} UEs, which describes a processing capability limitation of the AP and allows the system to achieve scalability. The K UEs are uniformly distributed over a square area of 1×1 km, and the distribution of the APs follows a hard core point process (HCPP)[5]. After the APs positioning, the coverage area is divided into J rectangle regions of the same size, each consisting of a CPU coordinating approximately L/J APs, where we have set $J = 4$. The values of L, N, U_{max}, and K vary and are specified throughout the results. In order to provide a better balance as to the amount of interference that affects each AP, we employ the wrap-around technique [14]. We focus on DL channels and consider $\tau_c = 200$ samples in each coherence block. The pre-log factor is set to $P_f = \tau_d/\tau_c$ for perfect knowledge of channel statistics, where $\tau_p = 10$, and $\tau_d = 190$. For imperfect knowledge, $P_f = 1 - (\tau_p/\tau_c) - \alpha$, where $\alpha = N_R \tau_p K/\tau_s \tau_c$. We consider that $B_s = 100$ MHz and $Ts = 0.5$ s, such that $\tau_s = 250000$ [18]. To calculate α and perform the correlation matrix estimation, we assume $N_R = 400$ and $N_{\mathbf{\Psi}} = 800$.

The total transmission powers of the UEs and APs are $p_i = 100\,\mathrm{mW}$, $\varrho_l = 1\,\mathrm{W}$, respectively. For computing the centralized power allocation (ϱ_k) we consider the following fractional power parameters $\upsilon = -0.5$ and $\kappa = 0.5$ [6]. We set the threshold of the LSFB algorithm to $\delta\% = 99.9$ and consider $\gamma = -50\,\mathrm{dB}$ in the proposed method. Moreover, we set $\Gamma\% = 98$, $\varepsilon = 0.02$ and $\zeta = \tau_p$ for the fine-tuning schemes. The parameters for EE are set as $\nu_l = 0.4$, $P_{\mathrm{tc},l} = 0.2$ W, $P_{0,l} = 0.825$ W, and $P_{\mathrm{ft},l} = 0.25$ W/(Gbit/s) [3]. We perform Monte-Carlo simulations to evaluate the system's performance in terms of average and cumulative distribution function (CDF) of the SE. We also evaluate the average numbers of APs connected to each UE ($|\mathcal{M}_k|$) and UEs per AP ($|\mathcal{U}_l|$).

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

The propagation model adopted is in accordance with the 3GPP line-of-sight (LOS)/non-line-of-sight-wireless (NLOS) Urban Micro (UMi) path-loss model defined in the Technical Report (TR) 38.901 [28]. We estimate the perfect correlation matrices $\mathbf{R}_k$ according to the local scattering spatial correlation model presented in [14, Sec. 2.6]. The parameter values used to set the entries for the UMi model and $\mathbf{R}_k$ can be found in Table I. We compare our solution with four other AP selection algorithms described earlier, which are the canonical CF [1], UCC [2], LSFB [3], and scalable CF [4].

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

Parameter	Value
Effective environment height, $\mathrm{h_E}$	1.0 m
Shadow fading standard deviation, σ_{SE}	4 dB
Antenna height AP, UE - $\mathrm{h_{AP}}$, $\mathrm{h_{UE}}$	11.65 m, 1.65 m
Rx noise figure (NF)	8 dB
Frequency center, bandwidth (B_s)	3.5 GHz, 100 MHz
Angular standard deviation (ASD)	20°
Antenna spacing	1/2 wavelength distance

Throughout the results, we modify the scalable CF AP selection method to consider the restriction U_{max}, described in Subsection II-B. For pilot allocation, we consider the algorithm presented in [4]. This is because we intend to compare our solution with the scalable CF scheme, which relies on this pilot assignment strategy to generate the AP clusters. This method assumes that the pilot assignment is performed by the AP with the strongest channel gain in the AP cluster of each UE. Moreover, the pilot assigned to each UE is the one that causes the least pilot contamination. The literature has introduced several strategies for pilot assignment in massive MIMO theory [29]–[31]. One can investigate how different AP selection schemes are affected by distinct pilot assignment strategies and vice-versa. However, a more profound discussion involving pilot allocation strategies is out of the scope of this paper.

We present the achievable SE results for different precoding choices, i.e., MR, LP-MMSE, P-RZF, and P-MMSE. It is noteworthy that MMSE precoding is a signal processing technique that maximizes the SINR for all UEs in the network based on channel estimates available on the CPU, efficiently suppressing interference among them. The distributed version of MMSE is called local MMSE (L-MMSE) precoding, which uses the local channel estimates available at each AP. These methods are not scalable since they need channel estimates of all UEs, but one can modify them to fulfill the scalability requirements. The key features of each scalable precoding method considered in this work are described below [6]:

- Distributed implementation
  - MR scheme: low-complexity precoding that maximizes the ratio between the received power and the square norm of the UL received combiner but cannot efficiently mitigate interference among UEs.
  - LP-MMSE scheme: an adaptation of the L-MMSE precoding that suppresses interference only from UEs served by the AP using locally available channel estimates, that is, with no cooperation among APs to this end. This precoding has higher complexity than the MR method but can provide better SE performance.
- Centralized implementation
  - P-MMSE scheme: the only difference between this precoding and the MMSE is that it does not consider all UEs in the uplink received combiner, but only the subset of UEs partially served by the same APs. The P-MMSE better mitigates interference than LP-MMSE, as the CPU entity has access to the channel estimates of several UEs, not only using local estimations.
  - P-RZF scheme: this technique simplifies the P-MMSE precoding by neglecting the estimation errors correlation matrix, allowing it to reduce time complexity. However, it still can suppress interference of the subset of UEs partially served by the same APs. It also performs better than LP-MMSE.

A. Cumulative Distribution Function

We start by evaluating the CDF of the achievable DL SE in a network consisting of $L = 100$ APs and $K = 20$ UEs. In Fig. 3, we compare the performance of the AP selection methods considering two scenarios and assuming perfect knowledge of channel statistics. The results compare APs that can serve a lower number of UEs ($U_{max} = 4$) with APs that can deal with more UEs ($U_{max} = 10$), which represent different processing capacity of their hardware.

TABLE II: Mean value and standard deviation (STD) of the number of APs connected to each UE for different AP selection methods.

AP selection method	$U_{max} = 4$		$U_{max} = 10$		Complexity
	Mean	STD	Mean	STD	
UCC (NS) [2]	20	0	50	0	$\mathcal{O}(LK \log K)$
LSFB (NS) [3]	30.28	1.46	30.28	1.46	$\mathcal{O}(L \log L)$
Scalable CF [4]	18.57	0.93	24.67	2.69	$\mathcal{O}(L\tau_p)$
Proposed method	19.94	0	49.94	0.57	$\mathcal{O}(LU_{max})$

In Fig. 3, the proposed solution can outperform the SE's of the remaining methods for the 95% likely UEs, when $U_{max} = 4$. It increases the SE of the 95% likely UEs by approximately 100% both for the P-MMSE and P-RZF compared to the scalable CF scheme, where the P-RZF achieves the same SE of the P-MMSE even though it is a less complex technique. In the distributed implementation, we observe a gain of about 163% using LP-MMSE and 75% with MR compared with [4], which corresponds to the expectations of both methods regarding interference mitigation. Additionally, the proposed method provides higher SEs with approximately the same number of APs connected per UE for $U_{max} = 4$, according to Table II. It is similar to UCC but with the advantage of being scalable and presenting less time complexity.

In Fig. 3, the increase in SE for the 95% likely UEs for $U_{max} = 4$ is related to the final AP cluster stage, which makes the worst UEs to connect to more APs. On the other hand, the 50% and 10% likely UEs benefit less from this stage, even if their SEs raise more. For example, an increase of 1

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2023.3235980

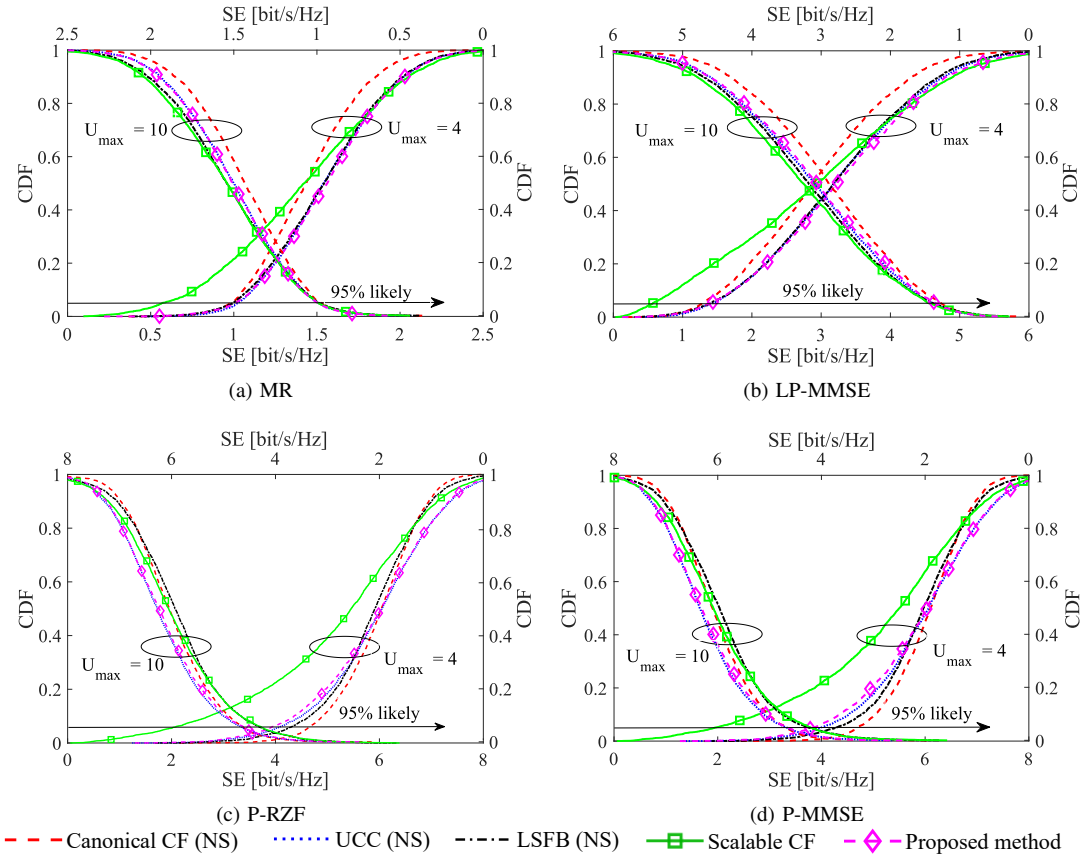


Fig. 3: Comparison of DL SE per UE of the proposed AP selection method with canonical CF [1], UCC [2], LSFB [3] and scalable CF [4]. Parameters setting: $U_{max} = 4$, $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

(bit/s/Hz) can substantially enhance the SE of the 95% likely UEs in percentage, while 1.2 (bit/s/Hz) will not present the same impact for the 50% and 10% likely UEs. Moreover, the intermediate AP cluster was crucial to improve performance. One can note that even though our solution makes the AP clusters comprise more APs, the average number of APs connected to each UE is not that different in Table II compared to the scalable CF method, since $U_{max} = 4$ is relatively small. Therefore, the matched-decision strategy enabled UEs and APs to establish more suitable connections, which helped in enhancing SE.

In Fig. 3, the gains offered by our method are not that impressive for $U_{max} = 10$, and some baseline solutions were slightly better in distributed implementation. As the APs can serve more UEs when $U_{max} = 10$, our solution makes the AP clusters even larger (as Table II shows), which generates more interference in the DL direction. Therefore, the proposed scheme has to be used jointly with more robust precoding techniques, such as P-MMSE and P-RZF, to provide gains in SE when $U_{max} = 10$. Otherwise, baseline solutions can present higher SEs since they serve the UEs using fewer APs, as Table II demonstrates. Nonetheless, the proposed method can present the lowest time complexity for $U_{max} < \tau_p$.

In Fig. 3, one can note that the proposed method allows the systems that employ APs with small U_{max} to provide SEs as high as those with a higher U_{max}. For instance, for the P-MMSE, the SE of the 50% likely UEs is about 6 (bit/s/Hz) and 6.3 (bit/s/Hz) for $U_{max} = 4$ and $U_{max} = 10$, respectively. These insights reveal that a network that employs APs serving many UEs does not necessarily provide a higher capacity. Furthermore, Fig. 3 also indicates that even if an AP can serve more UEs, it could reduce U_{max} through software to improve SE in some scenarios. Such results may also inspire future publications involving scalable UC networks by showing that U_{max} must not be too small or too large but properly suited to the network conditions. For instance, almost all AP selection schemes can outperform the canonical CF (at least in these controlled simulations) in Figs. 3 (a) and (b), emphasizing that an AP serving a higher number of UEs do not necessarily bring the highest SEs, as the more UEs the AP serves, the less is the allocated power and the higher is the interference.

B. Similarities with other AP Selection Methods

Recall that the proposed method can also provide scalability for other AP selection schemes. Fig. 4 compares the LSFB (NS) strategy with the MD LSFB described in Section III-C. Note that the MD LSFB strategy can achieve similar SEs as the LSFB with the advantage of being scalable. Moreover, it can perform as great as the scalable CF scheme. It can be noted in Fig. 3 that the LSFB (NS) scheme performs better than the scalable CF scheme for $U_{max} = 4$ and matches this one when $U_{max} = 10$. Hence, as the scalable version of the

LSFB does not present notable performance losses in Fig. 4, one can conclude that the MD LSFB presents SE levels as high as the scalable CF scheme.

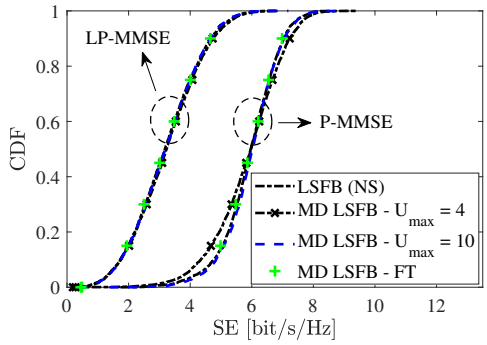


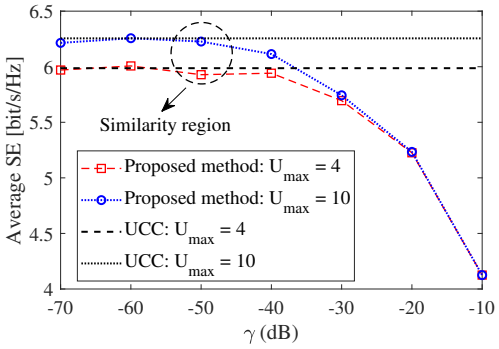
Fig. 4: Comparison of DL SE achieved when using LSFB and its scalable version, the MD LSFB. It is also presented a fine-tuned version of the MD LSFB (based on Algorithm 4) for $U_{max} = 10$, the MD LSFB-FT. Parameters setting: $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

A similar result can be observed for the UCC scheme since it is the method that most closely matches our scheme in Fig. 3. As Fig. 5 shows, the similarities between the proposed method and the UCC are related to the value of γ. For $\gamma < -40$ dB, the UE discovers a large number of APs in its vicinity, which makes the decisions of the APs predominant. Therefore, if all UEs choose many APs to connect, the APs will select those with the stronger channel gain in the intermediate stage. Thus, for $\gamma < -40$ dB, the proposed scheme is probably approaching a scalable version of UCC, which is called MD UCC. However, as γ increases, the UE decisions become even more restricted in the intermediate stage, i.e., the UE selects fewer APs in (20) and the similarities disappear. One can note that the similarity region (i.e., $\gamma < -40$ dB) allows the proposed method to reach high SEs while reducing EE. Therefore, the matched decision scheme must operate outside the similarity region to improve EE. That is, $\gamma > -40$ dB.

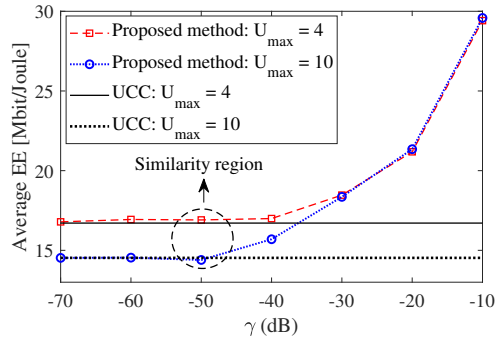
C. Performance of Fine-Tuning AP selection

Regarding the fine-tuning AP selection methods of Algorithms 3 and 4, one can observe in Table III that they substantially reduce the number of APs connected to each UE (especially for $U_{max} = 10$) while achieving similar results in terms of SE. These results demonstrate that fine-tuning AP selection schemes can potentiate the network performance, as the number of complex multiplications to precoding signals and estimate channels is proportional to $|\mathcal{M}_k|$ and $|\mathcal{U}_l|$. In Table III, one can note that the fine-tuning based on power allocation (Algorithm 3) can improve SE in LP-MMSE while reducing it in P-MMSE. As the distributed nature of this strategy aims to reduce the number of UEs connected in each AP ($|\mathcal{U}_l|$), it helps the interference mitigation of LP-MMSE. Therefore, it is more recommendable for distributed implementation.

On the other hand, both P-MMSE and LP-MMSE benefit from the fine-tuning based on SE (Algorithm 4) since this method reduces the number of APs connected to each UE by looking to the UE side. It is worth noting that these fine-tuning methods work in any AP selection scheme. For instance, Fig. 4 shows that we can keep the SE of the MD LSFB strategy almost unchanged while reducing the average number of APs per UE ($|\mathcal{M}_k|$) from 30.38 to 26.51 and the average number of UEs per AP ($|\mathcal{U}_l|$) from 6.05 to 5.3. In Fig. 4, the fine-tuning strategy based on SE is employed for $U_{max} = 10$, and the fine-tuned version of MD LSFB is called MD LSFB - FT.



(a) SE:P-MMSE



(b) EE:LP-MMSE

Fig. 5: Average DL SE and EE achieved by the proposed solution for different values of γ. Parameters setting: $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

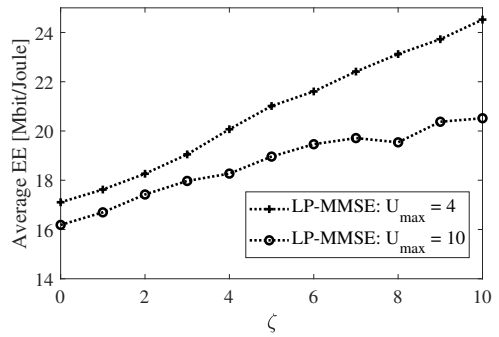


Fig. 6: Average EE achieved by the MD LSFB scheme after fine-tuning the AP clusters based on EE (Algorithm 5). Parameters setting: $L = 100$, $N = 1$ and $K = 20$. Perfect knowledge of channel statistics.

In Fig. 6, we analyze the impacts of Algorithm 5 on the system's EE. Specifically, it is evaluated the EE of the MD

TABLE III: Average number of APs per UE ($|\mathcal{M}_k|$), UEs per AP ($|\mathcal{U}_l|$) and DL SE for the proposed AP selection method before and after applying the fine-tuning methods from Algorithms 3 to 5. Parameters setting: $L = 100$, $N = 1$ and $K = 20$, $\Gamma\% = 98$, $\varepsilon = 0.02$, and $\zeta = \tau_p$. Perfect knowledge of channel statistics.

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

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

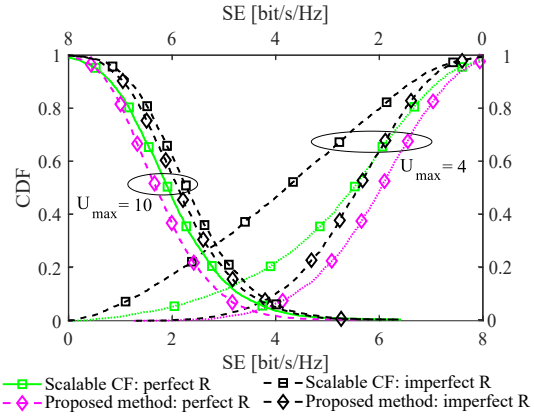


Fig. 7: Comparison of SE considering imperfect knowledge of channel statistics in the P-MMSE precoding. Parameters setting: $L = 100$, $N = 1$, and $K = 20$.

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

D. Impacts of Imperfect Knowledge of Channel Statistics on AP Selection Methods

The impacts of imperfect knowledge of the correlation matrices are illustrated in Fig. 7. It is possible to observe that the scalable CF scheme is the most affected by the imperfect $\widehat{\mathbf{\Psi}}_{tl}$ and $\widehat{\mathbf{R}}_{kl}$ when $U_{max} = 4$. In this scenario, the proposed method can offer gains of up to 315 % in the SE of the 95 % likely UEs compared to the scalable CF strategy. Nonetheless, both strategies are not greatly affected when $U_{max} = 10$.

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

E. Average Spectral Efficiency

From now on, we compare the performance of our proposed AP selection method with the only one that is also a scalable solution, the scalable CF scheme [4]. To compute the average SE, we consider only the LP-MMSE and P-MMSE since they provide the best interference mitigation for the distributed and centralized implementations, respectively. Besides, we are considering the perfect knowledge of channel statistics.

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

achieve the best SEs for LP-MMSE and P-MMSE, in which case one should set $U_{max} = 2$ and $U_{max} = 8$, respectively. In Fig. 8, one can also note that the scalable CF scheme needs to employ APs with greater values of U_{max} to achieve similar results than ours with smaller U_{max}.

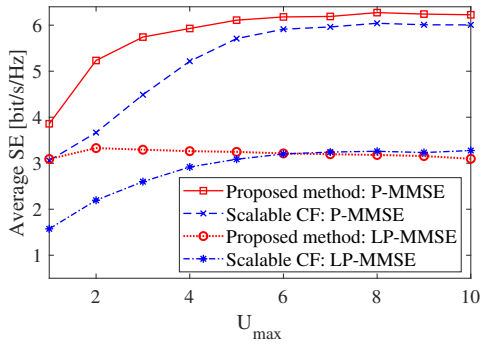


Fig. 8: Average DL SE versus U_{max}. Parameters setting: $L = 100$, $N = 1$, and $K = 20$.

Fig. 9 shows the average SE versus the number of antennas per AP, N. Note that the SE of our solution overcomes the scalable CF scheme when both use $U_{max} = 4$ and match the scalable CF when $U_{max} = 10$. However, the proposed method slightly loses performance for $N \geq 4$, when $U_{max} = 10$. The explanation for the results in Fig. 9 is similar to previous ones. That is, a suitable U_{max} allows the proposed solution to generate less interference and increase the SE, but an inappropriate U_{max} can degrade the SE and make baseline solutions perform slightly better.

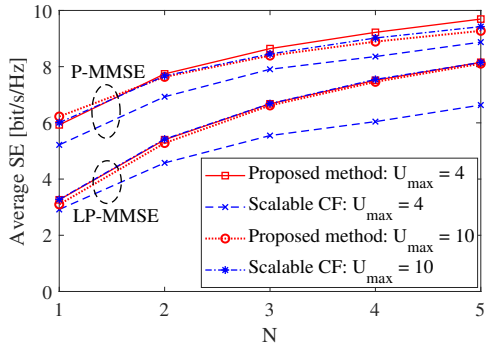


Fig. 9: Average SE as a function of the number of antennas per AP N. Parameters setting: $L = 100$ and $K = 20$.

Fig. 10 presents the analysis regarding achievable average SE versus the number of UEs, K. For LP-MMSE, the proposed method improves the average SE for all values of K by at most 17% compared to the scalable CF scheme with $U_{max} = 4$. Besides, our scheme performs better with $U_{max} = 4$ than $U_{max} = 10$, being outperformed by the scalable CF scheme with $U_{max} = 10$ and $K = 80$, but the difference is negligible. For P-MMSE, the proposed method improves the average SE up to 47% when $K = 70$ and U_{max} equal to 4. Moreover, one can note that $U_{max} = 10$ is more suitable for our solution for $K > 25$, i.e., when the massive

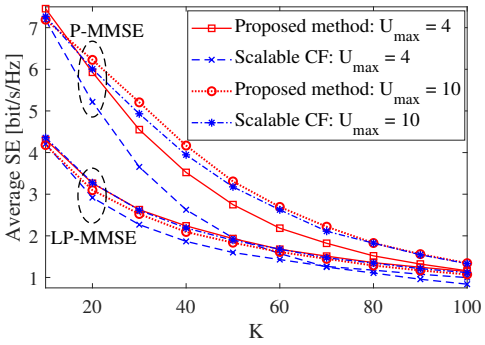
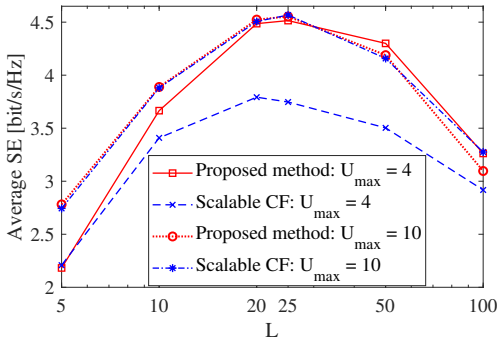
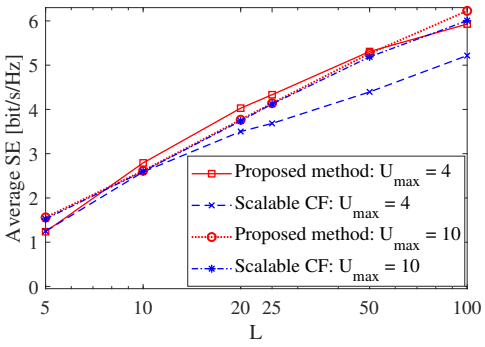


Fig. 10: Average SE versus the number of UEs K. Parameters setting: $L = 100$ and $N = 1$.



(a) LP-MMSE



(b) P-MMSE

Fig. 11: Average SE versus the number of APs L. Parameters setting: $K = 20$. The values of L and N varies in order to keep M fixed at 100, where $M = NL$.

MIMO condition ($M/K > 4$) is lost. This happens because the number of interfering UEs increases, while the number of APs in each AP cluster ($|\mathcal{M}_k|$) decreases. Thus, for $K > 25$ and $U_{max} = 4$, our solution could not provide gains compared to the scalable CF method with $U_{max} = 10$. This is because in addition to the clusters decrease, more UEs are partially served by the same APs, worsening the P-MMSE performance.

Fig. 11 presents the average SE versus the number of APs, L, where U_{max} is equal to 4 or 10, and N decreases as L increases to keep $M = 100$. One can note that for small values of L the SEs are not too high due to the low macro-diversity. However, as L increases and N decreases, the SE

improves, with the behavior of the SEs for $U_{max} = 4$ and $U_{max} = 10$ following the same explanation of previous results. In Fig. 11 (a), one can observe that the SE achieves a maximum value between $L = 20$ and $L = 25$. In this interval, the LP-MMSE could find the best balance between the strength of the received signal and interfering ones. However, for $L > 25$, the AP clusters are made up of more APs, consequently increasing interference and reducing the effectiveness of the LP-MMSE precoding. Additionally, Figs. 10 and 11 demonstrate that even if a system utilizes higher processing capacity APs, one can adapt U_{max} according to the network conditions to improve SE and reduce computational cost in some scenarios if it is used jointly with our solution.

V. Conclusions

We proposed a new scalable AP selection framework for UC CF massive MIMO systems. The algorithm is a competitive mechanism that exploits a matched-decision among the UEs and APs while guaranteeing connection for all UEs. The method consists of two stages, where the UEs first connect to an intermediate subset of APs and then form a final cluster. These steps aim to make the UEs and APs establish the best connection for both and then make the UEs expand their AP clusters intending to improve their SE. The method can improve the system's performance and afford scalability for baseline AP selection strategies. We also proposed three fine-tuning algorithms to be applied after the AP selection. The first two fine-tuning algorithms are based on allocated power and SE. Results indicate that they can enable UEs to achieve almost the same SE as before while reducing the number of APs serving each UE. The third one aims to improve the total EE. Results indicate that it can improve the EE up to 43% to the LP-MMSE. Besides, AP selection schemes and fine-tuning algorithms were evaluated under perfect and imperfect knowledge of channel statistics.

We analyzed the achievable SE in centralized and distributed network implementations by varying the numbers of UEs K, APs L and antennas N per AP. In our simulations, each AP can serve up to U_{max} UEs, and we evaluated networks whose APs could deal with different values for U_{max}. The results indicate that the proposed method can outperform baseline solutions and improve the SE of the worst UEs. For instance, our method can increase the SEs of the 95% likely UEs up to 163% and 100% in distributed and centralized implementations, respectively. The results also indicated that the scalable AP selection baseline solution requires that the APs serve more UEs than our solution to enable the network to achieve similar SEs. Additional results revealed that U_{max} has to be set appropriately to the network condition, e.g., number of UEs and network implementation. Therefore, although an AP can serve more UEs, it can reduce U_{max} to improve SE while reducing computational costs. Finally, it is noteworthy that our results are novel, useful for researchers working with AP selection methods, and can inspire new papers on the theme. Future works can expand our analyses to consider aspects such as non-reciprocity and limited fronthaul/backhaul capacity. Besides, it is worthwhile to optimize the number of UEs each AP can serve, and power allocation.

References

[1] 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.

[2] 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.

[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, Mar. 2018.

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

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

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

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

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

[9] 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.*, Dec. 2021, pp. 01–06.

[10] T. X. Vu, S. Chatzinotas, S. ShahbazPanahi, and B. Ottersten, "Joint power allocation and access point selection for cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–6.

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

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

[13] 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.

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

[15] S. Haghighatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8316–8332, Dec. 2017.

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

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

[18] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Mar. 2016, pp. 974–978.

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

[20] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, Jun. 2018.

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

[22] E. Bjornson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Signal Process*, vol. 59, no. 12, pp. 6086–6101, Aug. 2011.

[23] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends® in Communications and Information Theory*, vol. 9, no. 2–3, pp. 113–381, 2013. [Online]. Available: http://dx.doi.org/10.1561/0100000069

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

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

[26] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jul. 2019, pp. 1–6.

[27] 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.

[28] 3GPP, "5G; Study on channel model for frequencies from 0.5 to 100 GHz," *3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.901 Version 16.1.0 Release 16*, Nov. 2020.

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

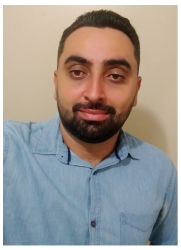
[30] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, Mar. 2018.

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

Marx Freitas received the B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of Pará (UFPA), Belém, Brazil, in 2018 and 2019, respectively, where he is currently pursuing the Ph.D. degree with the Postgraduate Program of Electrical Engineering (PPGEE). His research interests include massive MIMO and mobile transport networks for future wireless communication systems.

Daynara Dias Souza received a B.Sc. in electrical engineering in 2018 and an M.Sc. in electrical engineering with emphasis on telecommunication, in 2020, from the Federal University of Pará (UFPA), Belém, Brazil, where she is currently working toward a Ph.D. degree. She has experience in broadband communication systems. Her current research topic is resource allocation for wireless communication systems.

Dr. Daniel Benevides da Costa was born in Fortaleza, Ceará, Brazil, in 1981. He 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, Area: Telecommunications, from the University of Campinas, SP, Brazil, in 2006 and 2008, respectively. His Ph.D. thesis was awarded the Best Ph.D. Thesis in Electrical Engineering by the Brazilian Ministry of Education (CAPES) at the 2009 CAPES Thesis Contest. From 2008 to 2009, he was a Postdoctoral Research Fellow with INRS-EMT, University of Quebec, Montreal, QC, Canada. From 2010 to 2022, he was with the Federal University of Ceará, Brazil. From January 2019 to April 2019, he was Visiting Professor at Lappeenranta University of Technology (LUT), Finland, with financial support from Nokia Foundation. He was awarded with the prestigious Nokia Visiting Professor Grant. From May 2019 to August 2019, he was with King Abdullah University of Science and Technology (KAUST), Saudi Arabia, as a Visiting Faculty, and from September 2019 to November 2019, he was a Visiting Researcher at Istanbul Medipol University, Turkey. From 2021 to 2022, he was a Full Professor at the National Yunlin University of Science and Technology (YunTech), Taiwan. Since 2022, he is Principal Researcher of the AI and Digital Science Research Center at the Technology Innovation Institute (TII), a global research center and the applied pillar of Abu Dhabi's Advanced Technology Research Council. He is the Editor of several IEEE journals and has acted as Symposium/Track Co-Chair in numerous IEEE flagship conferences.



Gilvan Soares Borges received degrees as an engineer, M.Sc. and Ph.D. in electrical engineering from the Federal University of Pará (UFPA) in the years 2007, 2010 and 2016, respectively. He is currently a professor at the Federal Institute of Education, Science, and Technology of Pará (IFPA). Among his areas of interest are applied electromagnetism, computational intelligence, and broadband communication systems.

André Mendes Cavalcante received his Ph.D. in Electrical Engineering from the Federal University of Pará (UFPA), Brazil, in 2007. From 2007 to the beginning of 2016, he was a coworker at the Nokia Institute of Technology (INDT), Brazil, where he was engaged in several R&D projects related to wireless communications. Currently, he is a Senior Researcher at Ericsson Research, Brazil. His main research topic is mobile transport networks for future wireless communication systems.

Maria Valéria Marquezini received her Master of Science (Physics) degree in 1990 and Doctor of Science (Physics) degree in 1995 from the University of Campinas, Brazil. Between 1995-97 she worked as a post-doctoral researcher in the Materials Science Division of the Lawrence Berkeley National Laboratory (LBNL), conducting research activities in the field of electronic processes in semiconductors. From 1997 to 1999, she worked as an associate researcher at Physics Institute, UNICAMP. Since 2000, she has been working in the research department of Ericsson in Brazil and has been engaged in many different research projects in the telecommunications domain.

Igor Almeida (Member, IEEE) received the B.Sc. degree in computer engineering and the M.Sc. degree in electrical engineering from the Federal University of Pará (UFPA), Belém, Brazil, in 2010 and 2013, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering. He has been with Ericsson Research since 2016, involved with topics such as 5G, fronthaul networking, synchronization, and wireless communications.

Roberto Menezes Rodrigues received the Ph.D. degree in electrical engineering in 2012 from the Federal University of Pará (UFPA), Belém, Brazil. Currently, he is a professor of the Electrical and Biomedical Engineering Faculty at UFPA. He is also affiliated to the Applied Electromagnetism Laboratory (LEA). His current research interests are 5G networks, energy harvesting, and swarm robotics.

João C. Weyl Albuquerque Costa received a B.Sc. in electrical engineering from the Federal University of Pará (UFPA), Belém, Brazil, in 1981, an M.Sc. in electrical engineering from the Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, in 1989, and a Ph.D. 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.