Multi-Stream Allocation in Semi-Coherent Cell-Free MU-MIMO Systems

Esra Aycan Beyazıt[1], Mutlu Beyazıt[2], Andrei Belogaev[1], Jeroen Famaey[1], Miguel Camelo Botero[1]
[1]University of Antwerp - imec, IDLab, Antwerp, Belgium
{esra.aycanbeyazit, andrei.belogaev, jeroen.famaey, miguel.camelobotero}@uantwerpen.be
[2]University of Antwerp - Flanders Make, Antwerp, Belgium
mutlu.beyazit@uantwerpen.be

***Abstract*—This paper presents a novel stream allocation algorithm for semi-coherent Cell-Free Multi-User Multiple-Input Multiple-Output (CF-MU-MIMO) networks. In such networks, groups of phase-coherent Access Points (APs) are organized into clusters that operate coherently within themselves, while inter-cluster phase coherence is not maintained—reflecting practical limitations in achieving network-wide synchronization. The proposed algorithm is tailored for multi-cluster service, where each User Equipment (UE) can be simultaneously served by multiple clusters. To support this flexibility, stream paths are initialized and dynamically constructed based on the strongest links across all serving clusters. During the allocation process, both inter-cluster and inter-stream interference are mitigated using projection-based techniques, thereby improving spatial multiplexing efficiency. Compared to the greedy stream allocation which is commonly adopted in the literature as a low-complexity benchmark due to its near-optimal performance in downlink MIMO systems, the proposed method achieves higher performance with only a modest increase in complexity. Furthermore, significant enhancements in spectral efficiency is demonstrated relative to single-user-single-cluster allocation strategies in distributed cell-free systems.**

***Index Terms*—CF-MU-MIMO, stream selection, phase coherence, semi-coherent cell-free, and interference mitigation.**

I. Introduction

Traditional cellular architectures struggle to deliver consistent quality of service in dense and heterogeneous environments. To overcome these limitations, CF-MU-MIMO has been identified as a strong candidate for next-generation wireless networks, offering scalable and uniformly high performance by eliminating cell boundaries [1], [2]. In CF-MU-MIMO systems, numerous distributed APs jointly serve users over the same time-frequency resources, thereby enhancing macro-diversity and mitigating inter-cell interference. The architecture also enables more flexible user association and load balancing, which are key to supporting ultra-dense deployments and diverse traffic demands.

Achieving efficient resource coordination across the network, however, presents significant challenges. Full network-wide synchronization is practically infeasible due to hardware impairments, independent oscillators, and propagation-induced phase shifts [3]. As a result, phase misalignment among APs reduces the effectiveness of coherent joint transmission, limiting the theoretical performance gains. In addition to synchronization issues, Cell-Free Multiple-Input Multiple-Output (CF-MIMO) must address the constraints of limited-capacity fronthaul links and computational scalability, which affect the practical feasibility of centralized coordination and real-time Channel State Information (CSI) acquisition. Managing inter-user and inter-cluster interference while optimizing stream allocation under such constraints remains an open research area.

To overcome the aforementioned challenges, recent studies have proposed semi-coherent transmission schemes that integrate coherent transmission within locally phase-aligned AP clusters and non-coherent transmission across clusters [4], [5]. Despite their effectiveness, most existing works [4]–[8] consider UEs equipped with a single antenna, which limits their applicability to next-generation networks where multi-antenna UEs, receiving multiple streams at once, are increasingly common.

Effectively managing inter-stream interference in such non-coherent settings is crucial to fully exploit the potential of multi-stream transmission [9]. Stream allocation techniques based on Interference Alignment (IA) have been previously investigated in the context of conventional and heterogeneous cellular networks [10], [11]. However, these approaches were developed for cellular networks and often rely on rigid one-to-one user-to-cluster associations, making them too simplified for the distributed and flexible structure of CF-MU-MIMO systems. Our earlier work, submitted in [12], focused on one-to-one UE-cluster associations to explore the multi-stream transmission impact. Moving beyond the one-to-one setup, this paper extends the investigation to a more generalized and practical setting that aligns more closely with the CF-MIMO paradigm, where each user can be simultaneously served by multiple clusters, and each cluster can concurrently support multiple users. Such flexible association introduces additional complexity but also introduces new degrees of freedom for optimizing spectrum utilization and reducing energy consumption by enabling intelligent stream allocation that enables dynamical adaptation of traffic load, user location, and channel quality, thus enhancing spatial reuse and lowering unnecessary power expenditure.

We propose a novel stream selection framework that dy-

This research has been funded by the 6G-TWIN project, which has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the EU's Horizon Europe research and innovation program under Grant Agreement No 101136314.

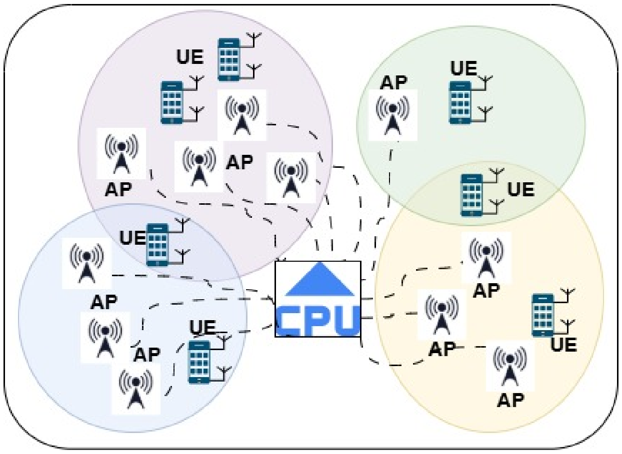

Fig. 1: Phase-coherent clusters composed of randomly distributed APs linked to CPUs serving to multi UEs

namically allocates streams across cluster-user pairs under semi-coherent transmission. The algorithm implements IA to mitigate inter-stream and inter-cluster interference while maximizing the network throughput. Streams are initialized using the strongest singular values of the user-cluster channel matrices, and the optimal stream path is determined by evaluating multiple stream sequences. Compared to greedy strategies, the proposed algorithm achieves substantial performance gains with only moderate complexity overhead, making it suitable for practical deployments in CF-MU-MIMO systems.

The remainder of the paper is organized as follows. Section II describes the system model and transmission framework for semi-coherent CF-MU-MIMO networks. Section III presents the AP clustering methodology and the proposed stream selection algorithm along with interference mitigation via IA. Section IV provides numerical results to evaluate the effectiveness and complexity of the proposed scheme. Finally, Section V concludes the paper.

II. System Model

The considered system model consists of a CF-MU-MIMO system with L APs, each with N_T transmit antennas, serving K UEs, each with N_R receive antennas. The set of UEs is denoted as $k \in \Gamma = \{1, ..., K\}$. All APs are connected to a Central Processing Unit (CPU) via fronthaul links as illustrated in Figure 1. Although a full phase-coherent model is desirable, achieving network-wide synchronization is generally impractical [3]. Therefore, the network is assumed to operate under a semi-coherent transmission model, where APs within each cluster are phase-aligned to enable coherent transmission, whereas inter-cluster transmissions remain non-coherent due to the absence of global synchronization. The system model for hybrid transmission can be defined as follows.

Let $\mathcal{L}_c$ denote the set of APs in cluster $c \in \{1, \dots, C\}$, and let $\mathcal{M}_k \subseteq \{1, \dots, C\}$ represent the set of clusters serving user k, referred to as the serving clusters in this paper. The cardinality $|\mathcal{M}_k|$ reflects the degree of transmission coherence.

Phase-aligned transmission within each cluster is achieved by grouping geographically proximate APs, which are tightly synchronized to form a virtual distributed MIMO array [13]. Each cluster can serve multiple UEs, and each UE can also be served by multiple clusters. The implemented phase-coherent AP clustering is explained in Section III-A.

The collective channel between user k and a serving cluster $c \in \mathcal{M}_k$, composed of the set of APs in $\mathcal{L}_c$, is denoted by $\mathbf{H}_{kc} \in \mathbb{C}^{N_R \times N_T |\mathcal{L}_c|}$ and is defined as:

$$\mathbf{H}_{kc} = \left[\alpha_{kl_1} \mathbf{h}_{kl_1} \quad \alpha_{kl_2} \mathbf{h}_{kl_2} \quad \cdots \quad \alpha_{kl_{|\mathcal{L}_c|}} \mathbf{h}_{kl_{|\mathcal{L}_c|}} \right], \tag{1}$$

where $l_1, l_2, \dots, l_{|\mathcal{L}_c|} \in \mathcal{L}_c$ and $\mathbf{h}_{kl} \in \mathbb{C}^{N_R \times N_T}$ represents the channel matrix from AP l to user k. Each element of $\mathbf{h}_{kl}$ follows an i.i.d. complex Gaussian distribution $\mathcal{CN}(0, 1)$, scaled by α_{kl} which is a large-scale fading including path loss and shadowing, and is computed per AP $l \in \mathcal{L}_c$ as follows.

$$\alpha_{kl} = 10^{-\mathrm{PL}(\mathrm{dis}_{kl})/10} \cdot 10^{-F_{kl}/10}, \tag{2}$$

where $\mathrm{PL}(\mathrm{dis}_{kl})$ is the path loss (in dB) as a function of the distance between AP l and user k, and F_{kl} denotes the shadow fading in dB.

The received signal at user k from cluster c includes both the desired signal and interference from other users served by the same cluster:

$$\mathbf{y}_{kc} = \mathbf{H}_{kc} \mathbf{x}_{kc} + \sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{m \in \mathcal{M}_j} \mathbf{H}_{km} \mathbf{x}_{jm}, \tag{3}$$

where the term $\mathbf{x}_{kc}$ represents the transmitted signal intended for user k from cluster c.

The total received signal at user k is then obtained by summing over all its serving clusters as follows.

$$\mathbf{y}_k = \sum_{c \in \mathcal{M}_k} \mathbf{y}_{kc} + \mathbf{n}_k, \tag{4}$$

where $\mathbf{n}_k \in \mathbb{C}^{N_R \times 1}$ is the Additive White Gaussian Noise (AWGN) vector, whose elements are independently and identically distributed as $\mathcal{CN}(0, \sigma^2)$.

The transmitted signal from each serving cluster $c \in \mathcal{M}_k$ to user k is denoted by $\mathbf{x}_{kc} \in \mathbb{C}^{N_T |\mathcal{L}_c| \times 1}$ and is defined as

$$\mathbf{x}_{kc} = \sqrt{P}\, \mathbf{T}_{kc} \mathbf{s}_{kc}, \tag{5}$$

where P is the transmit power allocated by cluster c for user k, and $\mathbf{T}_{kc} \in \mathbb{C}^{N_T |\mathcal{L}_c| \times q_{kc}}$ is the corresponding unitary precoding matrix. The number of streams allocated by cluster c to user k is denoted by q_{kc}, satisfying $q_{kc} \leq d_{kc}$, where $d_{kc} = \min(N_R, N_T |\mathcal{L}_c|)$. The associated symbol vector $\mathbf{s}_{kc} \in \mathbb{C}^{q_{kc} \times 1}$ is given by $\mathbf{s}_{kc} = [s_{kc,1}, \dots, s_{kc,q_{kc}}]^\mathsf{T}$ and satisfies the normalization condition $\mathbb{E}[\|\mathbf{s}_{kc}\|^2] = 1$. Equal power allocation among streams is assumed, i.e., $\mathbb{E}[|s_{kc,n}|^2] = 1/q_{kc}$ for $n = 1, \dots, q_{kc}$.

Furthermore, the maximum total number of streams in the network is calculated as follows.

$$r = \sum_{k=1}^{K} \sum_{c \in \mathcal{M}_k} d_{kc}, \tag{6}$$

The desired data symbols for user k are obtained by applying a postcoding matrix $\mathbf{D}_{kc} \in \mathbb{C}^{N_R \times q_{kc}}$ to the received signal $\mathbf{y}_{kc}$ from each serving cluster $c \in \mathcal{M}_k$. The post-processed signal corresponding to cluster c is given by

$$\hat{\mathbf{y}}_{kc} = \mathbf{D}_{kc}^H \mathbf{y}_{kc}. \tag{7}$$

The final decoded signal for user k is then obtained by summing the contributions from all associated clusters:

$$\hat{\mathbf{y}}_k = \sum_{c \in \mathcal{M}_k} \hat{\mathbf{y}}_{kc}. \tag{8}$$

The data rate for the i^{th} stream of user k is expressed as

$$R_{ki} = \log_2(1 + \gamma_{ki}), \tag{9}$$

where γ_{ki} is the SINR for the i^{th} stream of user k, given by

$$\gamma_{ki} = \frac{\sum\limits_{c \in \mathcal{M}_k} \frac{P}{q_k} \mathbf{d}_{kc}^{i\mathsf{H}} \mathbf{H}_{kc} \mathbf{t}_{kc}^{i} \mathbf{t}_{kc}^{i\mathsf{H}} \mathbf{H}_{kc}^{\mathsf{H}} \mathbf{d}_{kc}^{i}}{\mathbf{d}_{kc}^{i\mathsf{H}} \mathbf{B}_{ki} \mathbf{d}_{kc}^{i}}, \quad \forall k = 1, \ldots, K, \quad \forall i = 1, \ldots, q_k. \tag{10}$$

Here, $\mathbf{t}_{kc}^{i}$ is the i^{th} column of the precoding matrix $\mathbf{T}_{kc} \in \mathbb{C}^{N_T|\mathcal{L}_c| \times q_{kc}}$ for user k at cluster c, and $\mathbf{d}_{kc}^{i}$ is the corresponding column of the postcoding matrix $\mathbf{D}_{kc} \in \mathbb{C}^{N_R \times q_{kc}}$.

The interference-plus-noise covariance matrix $\mathbf{B}_{ki} \in \mathbb{C}^{N_R \times N_R}$ is given by

$$\begin{aligned} \mathbf{B}_{ki} = & \sum_{\substack{m=1 \\ m \neq i}}^{q_k} \sum_{c \in \mathcal{M}_k} \frac{P}{q_k} \mathbf{H}_{kc} \mathbf{t}_{kc}^{m} (\mathbf{t}_{kc}^{m})^{\mathsf{H}} \mathbf{H}_{kc}^{\mathsf{H}} \\ & + \sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{c \in \mathcal{M}_j} \sum_{q=1}^{q_j} \frac{P}{q_j} \mathbf{H}_{kc} \mathbf{t}_{jc}^{q} (\mathbf{t}_{jc}^{q})^{\mathsf{H}} \mathbf{H}_{kc}^{\mathsf{H}} + \sigma^2 \mathbf{I}_{N_R}. \end{aligned} \tag{11}$$

Finally, the total Sum Rate (SR) is calculated as

$$\text{SR} = \sum_{k=1}^{K} \sum_{i=1}^{q_k} \log_2(1 + \gamma_{ki}). \tag{12}$$

The main objective is to minimize interference while determining the optimal stream allocation for each UE and their serving clusters. In this context, the stream allocation scheme which maximizes the total sum rate of the network while guaranteeing at least one stream selection from each UE can be formulated as follows.

$$\{(\mathbf{T}_{kc}^*, \mathbf{D}_{kc}^*)\}_{\substack{k=1,\ldots,K \\ c \in \mathcal{M}_k}} = \underset{\{\mathbf{T}_{kc}, \mathbf{D}_{kc}\}}{\text{argmax}} \ \text{SR} \tag{13a}$$

$$\text{s.t.} \quad \sum_{c \in \mathcal{M}_k} q_{kc} \geq 1, \quad \forall k = 1, \ldots, K \tag{13b}$$

To address the challenge of phase misalignment in CF-MIMO networks, prior works have proposed clustering phase-coherent APs to enable coherent transmission within each cluster while allowing non-coherent transmission across clusters [4], [5], [9]. However, the joint problem of stream allocation across multiple serving clusters and users, with shared clusters serving multiple users, remains unexplored. In such settings, the resulting inter-cluster interference and intra-stream interference significantly affect system performance. Moreover, determining the optimal number of streams per user, under both spectral efficiency and sevice guarantee constraints, adds further complexity. To address these challenges, we propose a stream selection algorithm that dynamically allocates streams in a multi-cluster, multi-user serving scenario.

Alg. 1 Main Framework for Semi-Coherent Multi-Stream Selection

1: **Input:** Set of APs $\{1, \ldots, L\}$, users $\{1, \ldots, K\}$, and channel state information (CSI).
2: **Output:** Cluster configuration $\{\mathcal{L}_c\}$, selected streams per user, and associated beamforming vectors.
3: **Step 1: Cluster Formation**
Group phase-coherent APs into disjoint clusters $\{\mathcal{L}_c\}$ based on spatial proximity using a reference distance D_{ref}. (Adapted from [9])
4: **Step 2: System Setup**
Initialize system parameters and compute CSI matrices, $\mathbf{H}_{kc}$, $\forall k \in \{1, \ldots, K\}$, $\forall c \in \{1, \ldots, C\}$.
5: **Step 3: Stream Selection**
Invoke the stream selection algorithm (see Algorithm 2) to allocate spatial streams across clusters to each user, while mitigating inter-cluster interference via orthogonal projection.

III. The Proposed Framework

This work focuses on efficient multi-stream transmission in semi-coherent CF-MU-MIMO networks by introducing a flexible, interference-aware stream selection strategy tailored for multi-user, multi-cluster scenarios. The proposed algorithm supports multi-cluster transmission per user and allows clusters to serve multiple users simultaneously. To effectively manage inter-cluster and inter-stream interference in a semi-coherent system serving multiple users, we introduce a projection-based selection strategy that iteratively updates the stream pool by nulling mutual interference. The algorithm begins with phase coherent AP clustering [4], [9], followed by system initialization and CSI acquisition, and continues with the stream allocation phase, where the main performance gains are realized. The full framework is outlined in Algorithm 1.

A. Clustering Methodology

This sub-section explains a clustering algorithm that adopts a semi-coherent clustering strategy to address the challenges of phase misalignment in large-scale distributed CF-MIMO networks. Instead of assuming full phase coherence across all APs, the network is partitioned into non-overlapping clusters of APs that can achieve local phase alignment. Within each cluster c, APs share a common phase reference, enabling coherent joint transmission. However, transmissions across clusters are considered non-coherent due to independent local oscillators and unsynchronized timing references.

To address phase alignment challenges in distributed CF-MIMO systems, we adopt a semi-coherent clustering strategy, building on prior works such as [4], [9]. In this approach, APs are grouped into disjoint clusters to form sets $\mathcal{L}_c$. Each cluster $\mathcal{L}_c$ includes neighboring APs that are within a reference distance D_{ref}. This clustering ensures local phase coherence within each cluster. However, phase coherence across different clusters is not assumed, leading to a semi-coherent network architecture.

Clustering proceeds by first randomly selecting an AP and forming a cluster $\mathcal{L}_c$ by including all other unassigned APs within D_{ref}. The procedure continues iteratively until all APs are grouped into C non-overlapping clusters.

After cluster formation, user associations are established while selecting the streams that increase the sum-rates as explained in Section III-C. Each user $k \in \Gamma$ is served by a set of clusters, $\mathcal{M}_k$. This many-to-many association enables multiple users to be served by the same cluster and allows each user to receive service from multiple clusters, increasing the flexibility of stream assignment and spatial diversity.

This clustering methodology ensures that local phase alignment can be exploited for coherent transmission within each cluster, while non-coherent transmission across clusters accommodates the practical limitations of synchronization in large-scale deployments. The resulting semi-coherent transmission model forms the basis for the stream selection framework discussed in the next section. Further details are available in [9]. Inter-cluster interference is addressed in Section III-B.

B. Interference Mitigation

In semi-coherent CF-MIMO networks, both inter-cluster interference and inter-stream interference can significantly degrade performance. To address this, the stream selection method incorporates interference mitigation directly into the stream selection process. Specifically, each stream is selected such that it resides in the null space of the interference caused by previously selected streams. Hence, the newly added stream does not interfere with others already allocated, whether they originate from the same or different clusters. By projecting candidate streams onto the orthogonal subspace of previously selected channels, the algorithm preserves spatial orthogonality and suppresses inter-stream and inter-cluster interference, effectively.

Streams are extracted using the Singular Value Decomposition (SVD) of the effective channel between user k and each serving cluster $c \in \mathcal{M}_k$. Specifically, for each cluster-user pair (k, c), the SVD of $\alpha_{kc}\mathbf{H}_{kc}$ is given by $\mathbf{H}_{kc} = \mathbf{U}_{kc}\mathbf{S}_{kc}\mathbf{V}_{kc}^H$, where $\mathbf{U}_{kc} \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{V}_{kc} \in \mathbb{C}^{N_T|\mathcal{L}_c| \times N_T|\mathcal{L}_c|}$ are unitary matrices representing the receive and transmit beamforming directions at user k and cluster c, respectively. The diagonal matrix $\mathbf{S}_{kc}$ contains the singular values corresponding to stream strengths. The l^{th} columns of $\mathbf{U}_{kc}$ and $\mathbf{V}_{kc}$, denoted by $\mathbf{u}_{kc}^l$ and $\mathbf{v}_{kc}^l$, represent the receive and transmit beamformers for the l^{th} stream between user k and cluster c.

To manage interference during stream selection, IA is employed. Two types of inter-stream interference are considered:

- Interference *caused by* the selected stream (k, c, i) to other users' streams.
- Interference *experienced by* the selected stream (k, c, i) due to all other active streams in the network.

To suppress both interference types, virtual receiving and transmitting channels (VRC and VTC) are constructed [10]. After a stream is selected, the corresponding precoding and combining matrices for user k and cluster c are updated as

$$\mathbf{T}_{kc} = \left[\mathbf{v}_{kc}^1, \ldots, \mathbf{v}_{kc}^{q_{kc}}\right], \quad \mathbf{D}_{kc} = \left[\mathbf{u}_{kc}^1, \ldots, \mathbf{u}_{kc}^{q_{kc}}\right], \tag{14}$$

To ensure orthogonality concerning previously selected streams, the interfering channels are projected onto the null space of the selected stream's VRC and VTC. For any interfering user $j \neq k$ and cluster $c \in \mathcal{M}_j$, the projected channel is updated as $\mathbf{H}_{jc}^{\perp}$. The projection matrix is given by $\mathbf{P}_{\mathbf{x}}^{\perp} = \mathbf{I} - \frac{\mathbf{x}\mathbf{x}^H}{\|\mathbf{x}\|^2}$, where $\mathbf{x}$ denotes the beamforming vector of the selected stream. This projection ensures that future stream selections do not interfere with previously selected streams [10]. The IA procedure continues iteratively until all stream assignments are finalized, following the approach described in Algorithm 1 of [11].

C. Strongest-stream-initialized-path Comparative Stream Selection (S-CSS) Algorithm

In this section, we propose a recursive stream selection procedure tailored to a multi-cluster CF-MU-MIMO system. At each stream selection step, the algorithm incorporates the interference mitigation strategy described in Section III-B to manage both inter-cluster and inter-stream interference.

The algorithm constructs multiple stream paths, each initialized with the strongest stream received by a user from one of its potential serving clusters. For each user $k \in \Gamma$, the algorithm identifies the *strongest stream* among all candidate clusters $c \in \mathcal{M}_k$, based on the dominant singular value of the corresponding effective channel matrix $\mathbf{H}_{kc}$. The stream corresponding to this strongest singular value is selected to initialize a stream path for user k. This process is repeated for all users, resulting in the initial stream set Ω_0, such that $|\Omega_0| = K$.

Each stream in Ω_0 serves as the starting point of a stream path that is recursively extended through the proposed stream selection strategy.

Following initialization, the algorithm enters a recursive selection phase. At each iteration, it selects one stream (k, c, i) from the candidate pool Ω that maximizes the current total sum-rate. If no such stream provides an increase, the algorithm selects the stream that leads to the smallest decrease in sum-rate, with priority given to users who have not yet been assigned any streams. Selected streams are stored in the set Ψ, and removed from Ω. After each selection, IA-based orthogonal projections are applied to update the channel space and suppress interference between selected and remaining streams as described in Section III-B.

The process continues until no additional stream can be selected under the IA and rank constraints. The overall framework is presented in Algorithm 2.

IV. Performance Results

In this section, the performance of the proposed algorithm, S-CSS, is evaluated in a CF-MU-MIMO system setup where APs and UEs are independently and uniformly distributed in

Alg. 2 S-CSS Algorithm

Input: Set of users $\mathcal{K}$, clusters $\mathcal{C}$, CSI matrices $\{\mathbf{H}_{kc}\}$
Output: Beamformers $\{\mathbf{T}^*_{kc}, \mathbf{D}^*_{kc}\}$
Step 1. Compute SVDs and initialize candidate streams:
for each user $k \in \mathcal{K}$ and cluster $c \in \mathcal{M}_k$ **do**
  Compute SVD: $\mathbf{H}_{kc} = \mathbf{U}_{kc}\mathbf{S}_{kc}\mathbf{V}^{\mathsf{H}}_{kc}$
end for
  Initialize stream set:
  $\Omega = \{(k, c, l) \mid k = 1, \ldots, K;\ c \in \mathcal{M}_k;\ l = 1, \ldots, \text{rank}(\mathbf{H}_{kc})\}$
Step 2. Identify initial strongest streams:
Initialize $\Omega_0 = \emptyset$
for each user $k \in \mathcal{K}$ **do**
  Find strongest stream:
  $(k^*_k, c^*_k, l^*_k) = \arg\max_{k,c,l} \mathbf{S}_{kc}(l, l)$
  $\Omega_0 = \Omega_0 \cup \{(k^*, c^*, l^*)\}$
end for
Step 3. Initialize stream sequences:
for each initial stream $(k^*, c^*, l^*) \in \Omega_0$ **do**
  Initialize:
  $\Psi = \{(k^*, c^*, l^*)\}$; $i = 1$; $\Omega' = \Omega - \{(k^*, c^*, l^*)\}$
  while $|\Omega| \neq \emptyset$ **do**
    Compute SR_Ψ
    Select the strongest stream increasing the sum-rate
    $(k', c', l') = \arg\max_{k,c,l} \mathbf{S}_{kc}(l, l)$ and $\text{SR}_{\Psi \cup (k,c,l)} > \text{SR}_\Psi$
    if (k',c',l') does not exist and there is a user with no selected stream, select the strongest stream with minimum sum-rate decrease **then**
      $(k', c', l') = \arg\max_{k,c,l} \mathbf{S}_{kc}(l, l)$ and $\text{SR}_\Psi - \text{SR}_{\Psi \cup \{(k,c,l)\}}$
    end if
    if (k',c',l') does not exist **then**
      Break the loop
    end if
    $\Psi = \Psi \cup \{(k', c', l')\}$
    Perform IA with Algo. 1 in [11] and update $\mathbf{T}_{k'c'}$ and $\mathbf{D}_{k'c'}$
    $\Omega' = \Omega' - \{(k', c', l')\}$
  end while
  $\Xi_i = \Psi$ and $i = i + 1$
end for
Step 4. Select the best stream path:
$i^* = \arg\max_i \text{SR}_{\Xi_i}$
$\Psi^* = \Xi_i$
$\mathbf{T}^*_{kc}, \mathbf{D}^*_{kc} \Leftarrow$ beamformers corresponding to Ψ^*
Return: $\{\mathbf{T}^*_{kc}, \mathbf{D}^*_{kc}\}$ for all $k \in \mathcal{K}$, $c \in \mathcal{M}_k$

a 1×1km square area with a wrap-around topology. The minimum distance between each AP is 50m and $D_{ref} = 200$m. System parameters used in the simulations are listed in Table I.

To evaluate the performance of the proposed algorithm in CF-MU-MIMO networks, we adopt a greedy stream selection method as a baseline, specifically the algorithm described in Algorithm 2 of [11]. This method builds a single stream sequence by iteratively selecting the strongest stream, such as the one with the highest singular value, that increases the system sum rate. Although the greedy strategy is computationally efficient, its performance is limited because it evaluates only a single stream combination. In contrast, an exhaustive search considers all possible combinations and achieves optimal results. However, its complexity grows exponentially with system size, making it impractical for large networks, where even small scenarios become too time-consuming to evaluate.

In contrast, the proposed S-CSS algorithm provides an effective balance between performance and computational complexity. The complexity of stream selection algorithms is primarily determined by the number of calls to the IA

TABLE I: System Parameters

Parameter Name	Parameter Value
Transmit Power of APs	30dBm
Bandwidth	50MHz
Noise Power	-174dBm/Hz
Noise Figure	7dB
Simulation Area	1×1km
Path loss	$-30.5 - 36.7 \log_{10}\left(\frac{\text{dis}_{kl}}{1\text{ m}}\right)$dB
Shadowing std. dev.	4dB
Antenna number of each AP	4
Antenna number of each UE	2
Number of Monte Carlo Simulations	500

procedure, described in Section III-B and implemented in Algorithm 1 of [11]. Greedy stream selection constructs a single stream sequence and scales with r, the total number of stream candidates. In comparison, the S-CSS algorithm evaluates $K \times r$ combinations, where K is the number of users. For example, in a scenario with $L = 8$ and $K = 5$, the IA routine is invoked 10 times for greedy selection and 50 times for S-CSS. Despite this moderate increase, S-CSS significantly enhances performance by exploring multiple stream paths and remains far less complex than exhaustive search, making it a scalable solution for large CF-MU-MIMO systems.

We evaluated the performance of the proposed algorithm for a scenario where $L = 8$, $K = 5$. The results are provided in Figure 2. The proposed S-CSS algorithm is compared with greedy stream selection under the multi-cluster CF-MU-MIMO setup, as well as with the single-cluster-per-user model proposed in [12]. The sum-rate at the 90th percentile reaches 35 bps/Hz for S-CSS, compared to 30 bps/Hz for the greedy selection, indicating a 17% improvement. In addition, the proposed setup significantly improves performance over single-cluster models, where CSS [12] and greedy selection [11] achieve 30 bps/Hz and 25.5 bps/Hz, respectively. These results highlight the benefit of enabling multi-cluster service and show that S-CSS provides a favorable balance between performance and complexity in large-scale CF-MU-MIMO systems.

Moreover, the impact of phase-coherent clustering on performance is illustrated in Figure 3 for a scenario with $L = 16$ and $K = 8$. The sum-rate at the 90th percentile reaches 40.5 bps/Hz for the proposed S-CSS algorithm with coherent clustering, compared to 33 bps/Hz in the non-clustered system. Similarly, greedy stream selection benefits from clustering, with a sum-rate of 37 bps/Hz compared to 30 bps/Hz without clustering. These results highlight that phase-coherent clustering enhances the performance for both stream selection strategies, with S-CSS consistently providing the highest throughput across all configurations.

The average number of selected streams and the corresponding average sum-rate values in the small-scale CF-MU-MIMO scenario, where $L = 20$ and $K = 10$, are presented in Table II for different numbers of receive antennas at each UE. Unlike fixed stream allocation schemes, the proposed

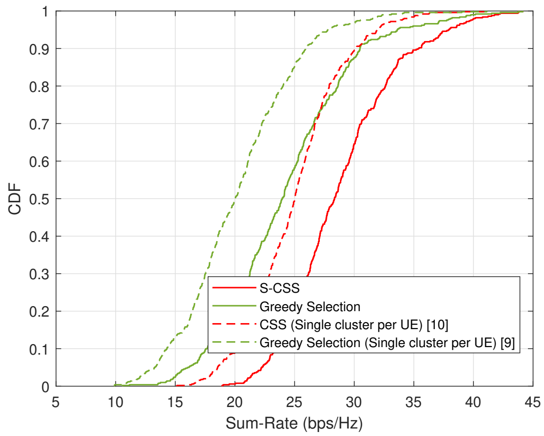

Fig. 2: Performance comparisons of S-CSS and greedy selection with both single-cluster-UE and multi-cluster–multi-UE system models for $L = 8$ and $K = 5$.

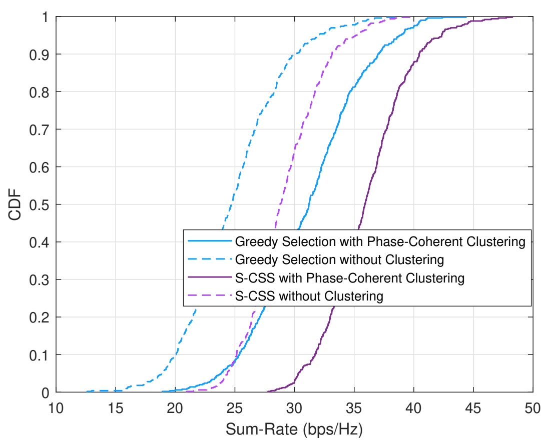

Fig. 3: Impact of coherent clustering on S-CSS and the greedy selection for $L = 16$ and $K = 8$.

algorithm dynamically selects the number of streams per user based on spatial channel characteristics and interference conditions. This flexibility enables more efficient use of spatial degrees of freedom.

V. Conclusion

This paper presented a flexible and scalable stream allocation framework for CF-MIMO systems with multi-antenna users and semi-coherent clustering. While previous works often assume single-antenna users and fixed user-cluster associations, our study considers a more practical setting in which each user can be served by multiple AP clusters and each cluster can support multiple users simultaneously. To address the resulting inter-cluster and inter-stream interference, we implement interference alignment-based stream selection algorithm that recursively updates the available stream pool by suppressing mutual interference through orthogonal projections. The algorithm integrates phase-coherent AP clustering with interference-aware stream allocation, offering an effective balance between performance and complexity. Numerical evaluations confirmed that the proposed approach significantly improves spectral efficiency compared to conventional greedy strategies. The proposed framework serves as a strong foundation for future distributed massive MIMO networks, with potential extensions such as mobility support and energy-efficient system design.

TABLE II: Average number of selected streams and sum-rate for different N_R values in a scenario with $L = 20$, $K = 10$.

Algorithm	$N_R = 2$		$N_R = 4$	
	Avg. Streams	**Sum-Rate**	**Avg. Streams**	**Sum-Rate**
S-CSS	10.16	37.43	12.12	49.75
Greedy Search	10.41	33.48	12.70	42.60

References

[1] E. Björnson and L. Sanguinetti, "Scalable cell-free massive mimo systems," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020.
[2] 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, 2021. [Online]. Available: http://dx.doi.org/10.1561/2000000109
[3] E. G. Larsson, "Massive synchrony in distributed antenna systems," *IEEE Transactions on Signal Processing*, 2024.
[4] G. Li, S. Wu, C. You, W. Zhang, and G. Shang, "Asynchronous cell-free massive mimo-ofdm: Mixed coherent and non-coherent transmissions," *IEEE Communications Letters*, 2024.
[5] L. Qin, R. Zhang, Y. Li, Y. Ruan, T. Li, D. Yang, H. Liu, and T. Yang, "A hybrid transmission scheme for cell-free massive mimo systems with phase offset," in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*. IEEE, 2024, pp. 1–6.
[6] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "Downlink resource allocation in multiuser cell-free mimo networks with user-centric clustering," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1482–1497, 2021.
[7] J. Zheng, J. Zhang, J. Cheng, V. C. Leung, D. W. K. Ng, and B. Ai, "Asynchronous cell-free massive mimo with rate-splitting," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1366–1382, 2023.
[8] S. Kurma, K. Singh, P. K. Sharma, C.-P. Li, and T. A. Tsiftsis, "On the performance analysis of full-duplex cell-free massive mimo with user mobility and imperfect csi," *IEEE Transactions on Communications*, 2024.
[9] U. K. Ganesan, T. T. Vu, and E. G. Larsson, "Cell-free massive mimo with multi-antenna users and phase misalignments: A novel partially coherent transmission framework," *IEEE Open Journal of the Communications Society*, 2024.
[10] M. Amara, M. Pischella, and D. Le Ruyet, "Enhanced stream selection for sum-rate maximization on the interference channel," *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, pp. 151–155, 2012.
[11] E. Aycan Beyazıt, B. Özbek, and D. Le Ruyet, "On stream selection for interference alignment in heterogeneous networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, pp. 1–18, 2016.
[12] E. A. Beyazıt, J. Famaey, N. Slamnik-Kriještorac, J. M. Marquez-Barja, and M. C. Botero, "Multi-stream transmission in cell-free mimo networks with coherent ap clustering," *IEEE Wireless Communications Letters*, 2025, submitted, March 2025.
[13] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.