

Fast Federated Reinforcement Learning-Based Load Balancing in Cell-Free Dynamic Scenarios

Karen A. Arcoverde, Vinicius N. H. Silva, Tadeu N. Ferreira, Roberto B. Di Renna

Abstract—The next-generation of mobile communications (6G) introduces the concept of cell-free (CF) architectures, where multiple distributed access points (APs) cooperate to jointly serve users, while still facing several challenges to be addressed. This paper evaluates AP congestion during high-density events and user equipment (UE) association by APs in uplink scenarios. The proposed FRL-M (Federated Reinforcement Learning Modified) scheme, combined with a new UE balancing technique across APs, increases Jain's fairness index and enhances fairness in resource distribution without degrading spectral efficiency (SE). The maximum improvement in Jain's fairness index reaches 91.18%, from LSF [1] to FRL-M, increasing from 0.395 to 0.755. In addition, FRL-M stabilizes approximately 2.8×10^3 steps earlier than standard Q-learning, reducing the fronthaul bottleneck and AP overload while accelerating learning convergence. These results demonstrate the potential of federated reinforcement learning-based strategies, combined with balanced UE allocation, to speed up convergence and enable fair resource management in Cell-Free massive Multiple-Input Multiple-Output (CFmMIMO) systems.

Keywords—6G, Cell-free massive MIMO, Federated reinforcement learning, Resource allocation.

I. INTRODUCTION

The main difference between Cell-Free massive Multiple-Input Multiple Output (CFmMIMO) and a cellular system is how users are associated with access points (APs). In a conventional cellular system, each user equipment (UE) is served by a single base station within a defined cell, and handovers occur when a UE moves between cells. In CFmMIMO, there are no cell boundaries, all distributed APs can jointly serve every UE, by coordinating transmissions to manage interference and share the service load across multiple APs [2].

Building on the CFmMIMO concept, federated learning (FL) can be applied to coordinate resource management across distributed access points without centralizing user data. In a CFmMIMO network, each AP locally processes information from the UEs it serves while computing updates to a shared learning model. These updates are sent to a central server (CPU), which aggregates them to improve the global model, then sends them back to the UEs to update their local models again. By keeping the raw data at the APs, federated learning preserves privacy and reduces the need for transmitting all user data to a central location [3].

Karen A. Arcoverde, Vinicius N. H. Silva, Tadeu N. Ferreira, Roberto B. Di Renna, Graduate Program in Electrical and Telecommunication Engineering, Fluminense Federal University, Niterói, Brazil, e-mail: {karenarcoverde, viniciusnhs, tadeu_ferreira, robertobrauer}@id.uff.br. This work was funded by CNPq grants 403784/2023-9, 403327/2025-3, and FAPERJ grant E-26/210.311/2024.

The article in [4] demonstrated the usefulness of Jain's fairness index as a metric for evaluating fairness in resource allocation, proposing an algorithm to balance the inherent trade-off between fairness and system efficiency in a single-channel TDMA system over slow fading channels. The authors in [5] incorporate adaptive mixed-resolution quantization into the FL process in a CFmMIMO, which significantly reduces communication overhead while maintaining model convergence. In [5], every AP serves all UEs in a network, without performing any selection of UE or AP. This approach increases the computational and fronthaul load at each AP, since signals and channel estimates are processed for every UE.

In contrast, the classical work [1] introduces a user-centric selection procedure where each UE is served only by a subset of APs. While this selection reduces the number of UEs each AP must handle, the framework does not explicitly limit the number of UEs that an AP can serve, which can still lead to increased AP overhead. In high-density scenarios such as stadiums, concerts, or large conferences, some APs may become heavily overloaded with many UEs while other APs remain underutilized. This creates an imbalance in resource utilization: heavily loaded APs suffer from fronthaul bottlenecks, while lightly loaded APs can operate with unconstrained fronthaul, resulting in overall uneven network performance.

Although the works [1], [6]–[12] have proposed AP selection strategies, they face the limitation that no mechanism exists to accelerate the learning process for identifying which APs/UEs correspond to the best or worst network performance metrics. Drawing inspiration in the QGradual technique in [13], we propose a solution called FRL-M (Federated Reinforcement Learning Modified) designed specifically for the CFmMIMO scenario, that aims to prevent AP underutilization and reduces fronthaul bottlenecks. Our scheme improves Jain's fairness index without degrading SE in dense scenarios, either slightly increasing it or keeping it unchanged, addressing a limitation observed in [5] and [1].

II. SYSTEM MODEL

This study considers a CFmMIMO uplink scenario with K single-antenna UEs and L APs with N antennas and distributed across the coverage area. The setup illustrated in Fig. 1 highlights the exchange of large-scale fading channel gains $\gamma_{l,k}$ between UEs and APs. At each AP, the FRL-M algorithm locally learns which UEs have the best or worst $\gamma_{l,k}$ and generates a Q-table $_{l,k}$ that is sent to the CPU, which performs selection based on [1] and load balancing. Each AP (working as the agent) trains locally, learning from its own environment, that is, only the UEs connected to the AP, without

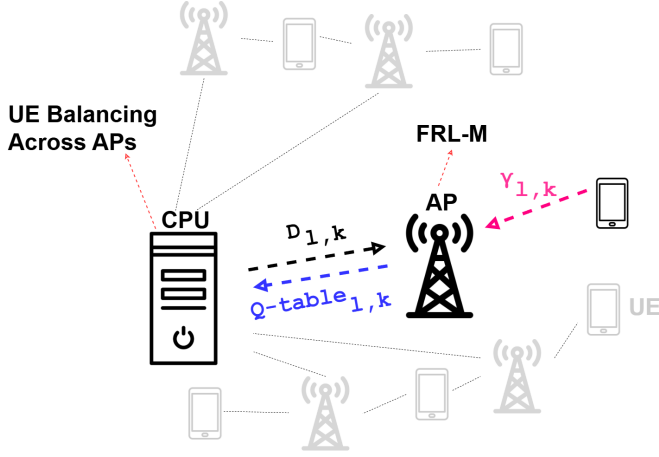


Fig. 1: CFmMIMO with distributed APs jointly serving UEs.

any knowledge of other APs. After this, the CPU returns the updated UE association matrix, $\mathbf{D}_{l,k}$, to the APs. The CPU acts as a federated coordinator, combining the local $Q\text{-table}_{l,k}$ from all AP agents to construct a global knowledge representation. This procedure is repeated over multiple federated rounds, in which APs continuously update their local $Q\text{-table}_{l,k}$ through Q-learning. Based on this representation, the CPU performs the global UE association and load-balancing decisions, and returns the updated association matrix $\mathbf{D}_{l,k}$ to the APs.

The algorithm in [1] performs the following selection to generate the matrix $\mathbf{D}_{l,k}$, where is assigned a Master AP, chosen as the one with the strongest large-scale fading (LSF) channel, and is allocated the pilot that experiences the least pilot contamination at that AP. Neighboring APs may serve the UE if they either have no UE on the same pilot or if the new UE has a better LSF [1] than their current UE. Unlike this approach, in our work the matrix $\mathbf{D}_{l,k}$ corresponds to the selection and also incorporates load balancing across the APs.

The SE experienced by UE k in uplink mode is shown below, following the definition in [1]:

$$\text{SE}_k^{(u,1)} = \frac{\tau_u}{\tau_c} \mathbb{E} \left\{ \log_2 (1 + \text{SINR}_k^{(u,1)}) \right\}, \quad (1)$$

where the pre-log factor $\frac{\tau_u}{\tau_c}$ accounts for the fraction of the coherence block dedicated to uplink data transmission, with τ_u denoting the number of uplink data symbols and τ_c the total coherence block length. The expectation $\mathbb{E} \{ \log_2 (1 + \text{SINR}_k^{(u,1)}) \}$ represents the achievable rate of UE k , averaged over the channel realizations. The SINR of each UE k in uplink mode [1] is described as follows:

$$\text{SINR}_k^{(u,1)} = \frac{p_k \left| \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_k \right|^2}{\sum_{i=1, i \neq k}^K p_i \left| \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_i \right|^2 + \mathbf{v}_k^H \mathbf{Z}_k \mathbf{v}_k}, \quad (2)$$

where the numerator of (2) represents the power of the desired signal received from UE k after combining at the APs. The denominator contains two contributions: the first term accounts for the interference from all other UEs $i \neq k$, and the second term represents the residual noise and estimation error. Here,

p_k is the transmission power of UE k and $\mathbf{v}_k$ is the combining vector used at the APs for UE k , which varies depending on the combining method, e.g., MR or P-MMSE [1]. As stated before, $\mathbf{D}_k$ is a selection matrix indicating which APs serve UE k , aggregating all the APs serving the same UE k from $\mathbf{D}_{l,k}$ into a single block-diagonal matrix. $\hat{\mathbf{h}}_k$ is the MMSE channel estimate between the APs and UE k , and $\mathbf{Z}_k$ represents the covariance of the residual interference and noise.

The large-scale fading channel gain in dB between AP l and UE k , normalized by the noise power is given by:

$$\gamma_{l,k} = G_0 - 10 \alpha \log(s_{l,k}) + \phi_{l,k} - N_0, \quad (3)$$

where G_0 is a reference constant representing the nominal gain at a reference distance. The second term, $10 \alpha \log(s_{l,k})$, models the path loss according to the distance $s_{l,k}$ between the AP l and UE k with path loss exponent α . The third term, $\phi_{l,k}$, accounts for shadow fading and captures slow variations of the channel due to obstacles in the environment, modeled as a Gaussian random variable. Finally, N_0 represents the total noise power in dBm over the system bandwidth, including thermal noise and the receiver noise figure.

III. PROPOSED MODIFIED FRL-QGRADUAL TECHNIQUE

This section describes the local learning capabilities of each AP through the FRL-M algorithm with a mechanism to associate UEs across the network. Each AP independently learns which UEs to serve, while the load balancing procedure ensures that heavily loaded APs are relieved by transferring users to underutilized APs. This approach preserves data privacy and reduces fronthaul bottleneck without degrading SE and improving fairness of the system in dense deployment scenarios. This section presents the Q-learning update rules, the integration of the FRL-M weight factor, and the detailed step-by-step algorithm implementing this procedure.

The Q-learning update at each access point l is defined to adjust the Q-values based on the observed reward and the estimated future reward. The idea is to allow each AP to learn which users to serve to maximize system performance. Thus,

$$Q_l(s, a) = Q_l(s, a) + \rho \left(R_l(s, a) + \sigma \max_{a' \in \mathcal{A}} Q_l(s', a') - Q_l(s, a) \right), \quad (4)$$

where $R_l(s, a)$ is the reward corresponding to the large-scale fading channel gain between AP l and UE k . Here, s and s' represent the current and next AP (state), respectively, and a and a' correspond to the actions of selecting a UE, where a is the current action and a' is the next action, e.g., selecting the UE with the highest reward. $\mathcal{A}$ is the set of possible actions, σ is the discount factor, and ρ is the learning rate.

The FRL-M method emerges as an alternative to accelerate this learning process, representing a modification of the technique proposed in [13]. In our approach, besides applying it in a CFmMIMO environment, we input a weight factor w_t applied to the Q-learning update and integrate FL. The FRL-M for each AP l with the weight factor w_t is given by:

$$Q_l(s, a) = Q_l(s, a) + w_t \cdot \Gamma \quad (5)$$

where $\Gamma = \rho (R_l(s, a) + \sigma \max_{a' \in \mathcal{A}} Q_l(s', a') - Q_l(s, a))$, and

$$w_t = \begin{cases} \log(\sqrt{n}) + 1 - \frac{t(\log(\sqrt{n}))}{\delta}, & \text{if } 0 < t < t_\delta, \\ 1, & \text{if } t \geq t_\delta, \end{cases} \quad (6)$$

The weight w_t is initialized based on the number of agents n , which act as APs, ensuring that $w_t > 1$ at the beginning. It gradually decays to 1 over the communication rounds t . A predefined constant δ determines the speed at which w_t decays to 1, and t_δ represents the time step at which w_t reaches 1 [13].

The selection builds upon the approach in [1], incorporating a dynamic evaluation of channel conditions, AP load, and potential connection associations to adapt the serving strategy in real time. This mechanism enhances the robustness of the network for sudden changes in user distribution and contributes to faster convergence of the learning process. The following steps describe the procedure for balancing UEs among APs:

- 1) **UE selection for associating connections:** For each AP, connected UEs that have not been associated are identified. The UE with the worst channel gain at that AP is selected as a candidate for relocation. Its gains across all other APs are evaluated and sorted to identify the best alternative APs.
- 2) **Criteria and execution for connection associations:** A UE is associated only if the percentage difference in load between the APs is at least $N_{th}\%$ and the current AP is more heavily loaded. The association matrix $\mathbf{D}_{l,k}$ is updated, the UE counts per AP are recalculated, and the UE is marked as associated. This process repeats until no further associations are possible.

The proposed approach is shown in Algorithm 1, which integrates the FRL-M with the AP load balancing procedure, based on the concept introduced in [1], to efficiently redistribute UE connections across the network.

IV. NUMERICAL RESULTS

The proposed FRL-M approach is evaluated from three complementary perspectives. First, the convergence analysis investigates how fast the agents learn the environment rewards compared with standard Q-Learning. Then, the user density scenarios analyze the impact of increasingly non-uniform UE distributions on AP load balancing and Jain's fairness index. Finally, the spectral efficiency analysis evaluates whether the proposed balancing strategy improves fairness without degrading the SE, considering MR combining scheme and comparing the performance of the congested first 4 APs with the remaining APs.

A. Convergence analysis

The simulation parameters in Table I follow [1] in CFm-MIMO uplink mode, with UEs assumed to be independently

Algorithm 1: FRL-M with UE Balancing Across APs

```

1 Each AP trains its local Q-tablel,k;
2 Each AP sends its Q-tablel,k to the CPU;
3 Input: Q-tablel,k;
4 Performs selection based on [1] and generate  $\mathbf{D}_{l,k}$ ;
5 Initialize the set of associated UEs:
   associated_UEs  $\leftarrow \{\}$ ;
6 for each AP  $ap = 0 \dots L - 1$  do
7   while True do
8     Identify UEs connected to  $ap$  that have not
       been associated yet;
9     if no such UEs exist then
10      break;
11     Select the UE with the worst gain at this AP;
12     Obtain the gains of this UE at all other APs.
       Sort them in descending order excluding  $ap$ ;
13     for each candidate AP do
14       Compute the percentage difference in UE
         counts between the current AP and the
         candidate AP;
15       if difference  $\geq N_{th}\%$  and current AP has
         more UEs then
16         Associate the  $UE_{\text{selected}}$  from current
           AP to candidate AP;
17         associated_UEs = associated_UEs +
            $UE_{\text{selected}}$ ;
18         Update the UE counts per AP;
19         Update  $\mathbf{D}_{l,k}$  to reflect the association
           connections;
20         break;
21     if no UE was associated then
22      break;
23 Output: CPU returns the association matrix  $\mathbf{D}_{l,k}$  to
    each AP;
    
```

TABLE I: Simulation parameters for the cell-free setup.

Parameter	Value
Number of UEs	$K = 100$
Number of APs	$L = 400$
Number of antennas per AP	$N = 1$
Transmit power per UE	$p_k = 100$ mW
Number of uplink symbols	$\tau_u = 190$
Coherence block length	$\tau_c = 200$
Number of pilot symbols	$\tau_p = 10$
Number of realizations	1000
Bandwidth	$B = 20$ MHz
Pathloss exponent	$\alpha = 3.76$
Noise figure	7 dB
Uplink noise power	$N_0 = -94$ dBm
Reference path gain	$G_0 = -35.3$ dB
Discount factor of Q-learning	$\sigma = 0.9$
Learning rate of Q-learning	$\rho = 0.2$
Time step of Q-learning	$t_\delta = 20000$
AP load imbalance threshold	$N_{th} = 50$

and uniformly distributed over a 2×2 km area. Additionally, the simulation begins with τ_p UEs assigned distinct pilots,

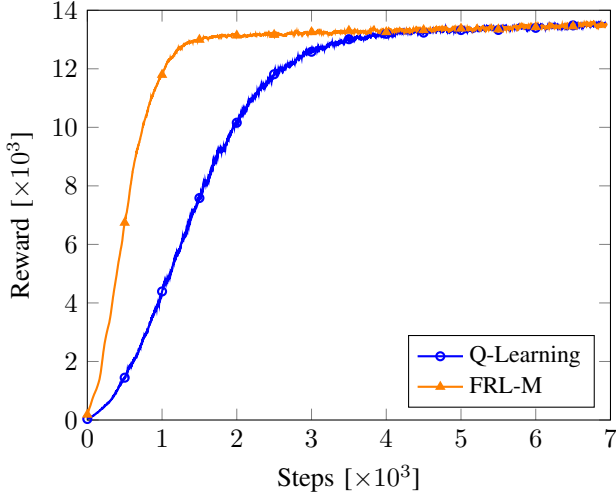


Fig. 2: Learning curves as a function of steps.

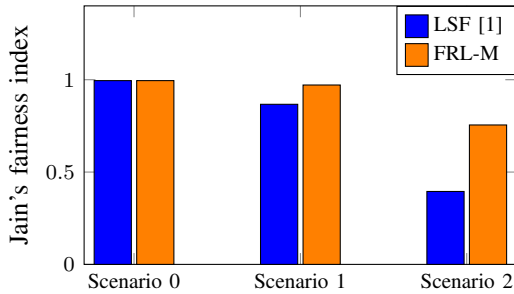


Fig. 3: Comparison of Jain's fairness index for each scenario.

after which the remaining UEs are connected sequentially. Furthermore, spatially correlated fading is considered, with APs deployed 10 m above the UEs. These settings are then used to generate the results shown in Fig. 2.

The learning curve in Fig. 2 represents the reward values of all agents over the episodes. It illustrates the comparison between the traditional Q-Learning algorithm and FRL-M. It is evident that FRL-M converges significantly faster, reaching high reward levels in fewer episodes, whereas Q-Learning requires a larger number of steps to achieve similar values. Specifically, FRL-M stabilizes approximately 2.8×10^3 steps earlier than standard Q-Learning. Thus, agents using FRL-M were able to learn more efficiently about the environment's reward, reflecting an accelerated collective learning process.

B. User density scenarios

Using the same parameters from Table I, with 200 additional UEs, resulting in three scenarios that investigate the impact of user density. A 200×200 m sub-square is placed at the center of the 2×2 km area, containing 4 APs. Other APs are located outside the sub-square. This configuration is maintained for all the following scenarios, changing only the UEs distribution, as illustrated in Fig. 4:

- **Scenario 0:** All UEs are uniformly distributed across a 2×2 km square.
- **Scenario 1:** 50% of the UEs are concentrated inside the sub-square. The remaining 50% are distributed outside.

- **Scenario 2:** 90% of the UEs are concentrated inside the sub-square, and only 10% outside.

As shown in Fig. 3 compares Jain's fairness index between the LSF [1] and FRL-M across the three scenarios. As the UE distribution becomes less uniform, the gain in the fairness index increases, the algorithm achieves a more equitable allocation of resources. The maximum improvement reaches 91.18%, from LSF [1] to FRL-M, increasing from 0.395 to 0.755, indicating that the proposed algorithm effectively balances the load among APs. This demonstrates the capability of the approach to enhance fairness in resource allocation in order to reduce the fronthaul bottleneck, especially in scenarios with high UE congestion at an AP.

C. Spectral efficiency

The CDF plots in Fig. 5, when comparing the traditional selection method with FRL-M, show that, for UEs are uniformly distributed, there is no loss or gain in SE, while the more uneven the UE distribution, the higher the SE gain. Furthermore, the first 4 APs achieve a larger SE gain than other APs. Other APs, which receive UEs offloaded from the congested first 4 APs, experience only a small SE gain or a slight SE loss, reflecting the combined effect of user redistribution and interference management. Overall, balancing UEs across APs does not degrade SE; in fact, SE either slightly increases or remains unchanged compared to the LSF [1]. This demonstrates that, despite achieving a significant gain in Jain's fairness index, there was no SE degradation in the scenarios studied.

ACKNOWLEDGEMENT

This work was funded by CNPq grants 403784/2023-9, 403327/2025-3, FAPERJ grant E-26/210.311/2024.

V. CONCLUSIONS

This paper proposes a fast FL method for load balancing in Cell-Free networks under dense scenarios. The proposed approach achieves faster learning convergence than standard Q-learning to identify which UEs present the best and worst LSF conditions [1], stabilizing approximately 2.8×10^3 steps earlier. Moreover, FRL-M improves Jain's fairness index without degrading SE, alleviates fronthaul bottlenecks and AP overload, and enhances fairness in resource distribution. The maximum improvement in Jain's fairness index reaches 91.18%, increasing from 0.395 to 0.755, highlighting the effectiveness of the proposed load balancing strategy.

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] S. Elhoushy, M. Ibrahim, and W. Hamouda, "Cell-Free Massive MIMO: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 492–523, 2022.
- [3] T. T. Vu, D. T. Ngo, N. H. Tran, H. Q. Ngo, M. N. Dao, and R. H. Middleton, "Cell-Free Massive MIMO for Wireless Federated Learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6377–6392, 2020.

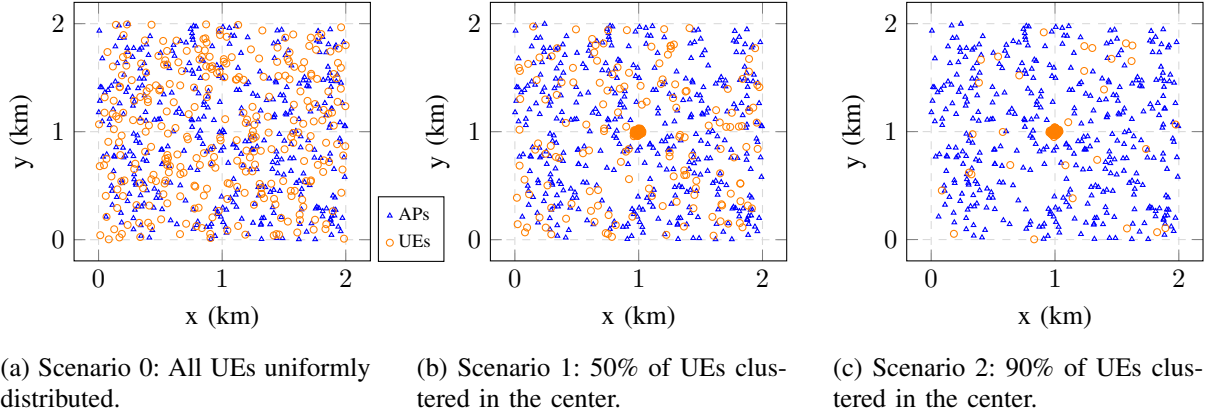


Fig. 4: Geometrical distribution of APs and UEs for all three scenarios. Each scenario depicts a different UE density scenario.

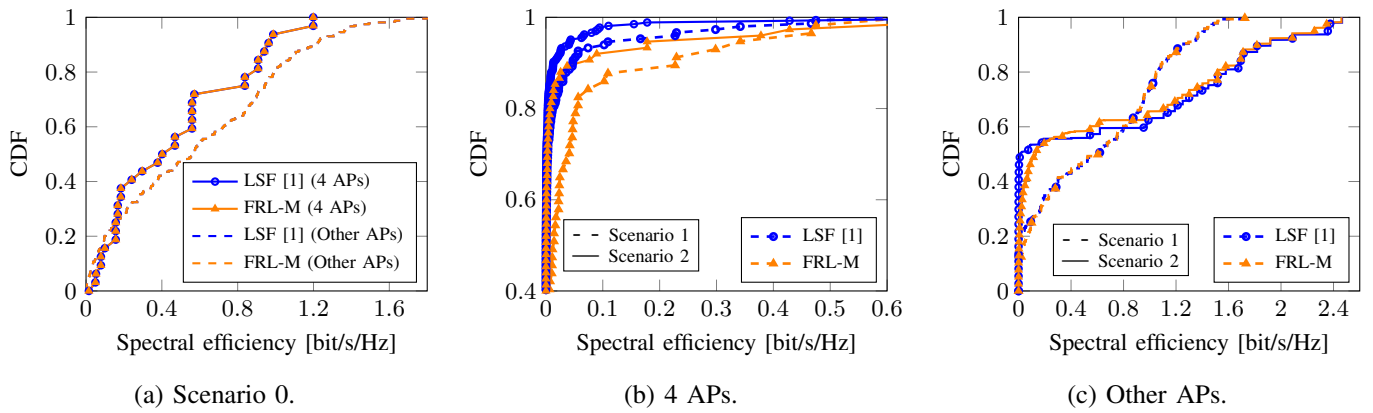


Fig. 5: CDF of the spectral efficiency for the three scenarios, comparing LSF [1] and FRL-M for the first 4 APs and Other APs.

- [4] C. Guo, M. Sheng, Y. Zhang, and X. Wang, "A Jain's Index Perspective on α -Fairness Resource Allocation over Slow Fading Channels," *IEEE Communications Letters*, vol. 17, no. 4, pp. 705–708, 2013.
- [5] A. Mahmoudi and E. Björnson, "Adaptive Quantization Resolution and Power Control for Federated Learning over Cell-free Networks," in *2024 IEEE Globecom Workshops (GC Wkshps)*, 2024, pp. 1–6.
- [6] E. Björnson and L. Sanguinetti, "Making Cell-Free Massive MIMO Competitive With MMSE Processing and Centralized Implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2020.
- [7] C. Wei, K. Xu, X. Xia, Q. Su, M. Shen, W. Xie, and C. Li, "User-Centric Access Point Selection in Cell-Free Massive MIMO Systems: A Game-Theoretic Approach," *IEEE Communications Letters*, vol. 26, no. 9, pp. 2225–2229, 2022.
- [8] N. Ghiasi, S. Mashhadi, S. Farahmand, S. M. Razavizadeh, and I. Lee, "Energy Efficient AP Selection for Cell-Free Massive MIMO Systems: Deep Reinforcement Learning Approach," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 1, pp. 29–41, 2023.
- [9] Z. Gao, Q. Zhang, J. Liu, Z. Du, and Y. Li, "DRL-Based AP Selection in Downlink Cell-Free Massive MIMO Network With Pilot Contamination," *IEEE Communications Letters*, vol. 28, no. 6, pp. 1432–1436, 2024.
- [10] 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, 2020.
- [11] M. S. Abbas, Z. Mobini, H. Q. Ngo, H. Shin, and M. Matthaiou, "Joint AP Selection and Power Allocation for Unicast–Multicast Cell-Free Massive MIMO," *IEEE Internet of Things Journal*, vol. 12, no. 22, pp. 47 135–47 150, 2025.
- [12] H. Cao, H. Zeng, X. Zhu, Q. Yang, and N. Wu, "Adaptive Fuzzy Logic Based AP Selection for Scalable User-Centric Cell-Free Massive MIMO," *IEEE Trans. Vehicular Technology*, vol. 74, no. 2, pp. 3429–3433, 2025.
- [13] Y. Pang, Z. Ni, and X. Zhong, "A fast federated reinforcement learning approach with phased weight-adjustment technique," *Neurocomputing*, vol. 626, p. 129550, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S092523122500222X>