Optimization of the Scheduling Problem in Cell-Free Massive MIMO Communication Systems

Marta Benavente Vilas, Óscar Fresnedo Árias, and Dariel Pereira Ruisánchez

Faculty of Computer Science, Universidade da Coruña, 15071 A Coruña, Spain
Centro de Investigación CITIC, Universidade da Coruña, 15071 A Coruña, Spain
Correspondence: marta.benavente.vilas@gmail.com

DOI: *https://doi.org/10.17979/spu.23.c25*

Abstract: This work addresses the selection of subsets of access points (APs) to serve users cooperatively in Cell-Free Massive MIMO (Multiple-Input Multiple-Output) systems. Traditional cellular architectures suffer from coverage and interference issues at cell edges. In contrast, Cell-Free networks distribute simpler APs uniformly across the area, enhancing spectral efficiency and connection quality. We simulate realistic communication scenarios to train a Deep Contextual Bandits (DCB) model that optimizes AP assignment based on user channel gains and interference. Alternative data-driven approaches, such as loss-based models and fuzzy clustering, are also developed and evaluated. Results show that DCB offers scalable, adaptive performance for next-generation wireless networks like B5G and 6G.

1 Introduction

Cell-free massive multiple-input multiple-output (CF-mMIMO) systems have emerged as a promising alternative to traditional cellular architectures. Instead of relying on fixed cell boundaries and a small number of powerful base stations, CF-mMIMO distributes a large number of simpler access points (APs) across a coverage area. These APs cooperate to jointly serve user equipments (UEs), effectively removing the cell-edge effect that limits performance in conventional systems. This decentralized architecture improves coverage, enhances spectral efficiency, ensures a more uniform quality of service (QoS), and adapts more flexibly to dynamic environments driven by user mobility Ngo (2020); Ngo et al. (2017b).

In CF-mMIMO, each UE can be simultaneously served by multiple APs working in cooperation, which significantly boosts coverage and capacity while delivering a consistent user experience, independent of geographic location. Unlike cellular systems where each base station carries a large number of antennas, CF-mMIMO networks rely on many distributed APs, each equipped with only a few antennas. This structural difference mitigates interference, reduces signal degradation, and enhances spectral efficiency across the entire network.

Alongside APs and UEs, a third key component plays a central role: one or more centralized processing units (CPUs) coordinate signal processing, resource allocation, and inter-AP cooperation. The fundamental principle of CF-mMIMO is thus dynamic AP cooperation, designed to maximize per-user QoS and ensure efficient distribution of resources in dense, high-mobility scenarios.

However, this paradigm also raises new challenges. Scheduling remains a fundamental problem in wireless network management. The goal is to maximize spectral efficiency while controlling interference, thereby improving overall system performance without sacrificing fairness or

user QoS. The task becomes especially difficult in CF-mMIMO due to the need for constant coordination among APs, the dynamic mobility of UEs, and the combinatorial complexity of maximizing the global sum-rate under strict latency requirements. These challenges motivate the exploration of novel solutions that combine scalability, adaptability, and efficiency, moving beyond traditional heuristics toward learning-based approaches capable of meeting the demands of next-generation wireless systems.

Several research efforts have addressed the resource allocation problem in wireless networks, with particular attention to CF-mMIMO systems. Early approaches relied on classical heuristics such as round-robin scheduling, greedy user selection, or sequential allocation based on the strongest channel gains Buzzi et al. (2017); Ngo et al. (2017a). While these strategies offer low computational complexity and fast assignments, they suffer from poor adaptability and limited spectral efficiency, especially in highly dynamic or dense user environments.

Deep reinforcement learning (DRL) has also been explored for wireless resource allocation, employing algorithms such as Deep Q-Networks (DQN) or Actor–Critic methods to optimize scheduling policies through interaction and experience Challita et al. (2019); Liang et al. (2019). Although highly adaptive, these solutions require significant computational resources and long exploration phases before converging to optimal performance, which makes them less practical for latency-sensitive scenarios.

More recently, graph neural networks (GNNs) have been proposed to capture the spatial and interference relationships among UEs and APs in CF-mMIMO Pereira-Ruisánchez et al. (2025). These models have demonstrated improvements in spectral efficiency compared to traditional methods. However, their reliance on large training datasets and potential difficulties in generalizing across dynamic environments limit their applicability.

In contrast, Contextual Bandits have received relatively little attention in the wireless networking domain. Their ability to make fast, context-aware decisions without explicitly modeling long-term dynamics makes them particularly attractive for resource allocation tasks. This gap in the literature reinforces the motivation for our work, which introduces Deep Contextual Bandits (DCB) as a central strategy for CF-mMIMO scheduling.

Finally, it is worth noting that few studies have investigated fuzzy clustering or loss-based models for AP assignment, even though these approaches can provide complementary perspectives by balancing interpretability and optimization-driven design.

2 System model

We consider a cell-free massive MIMO (CF-mMIMO) system composed of M distributed access points (APs), each equipped with a small number of antennas, and K user equipments (UEs). A set of one or more centralized processing units (CPUs) coordinates signal processing, resource allocation, and inter-AP cooperation. Unlike conventional cellular architectures with fixed cell boundaries, in CF-mMIMO every UE can be simultaneously served by multiple APs, which dynamically cooperate to improve the perceived quality of service (QoS). This architecture eliminates the cell-edge effect, ensures more uniform service across the coverage area, and adapts naturally to dynamic environments with mobile UEs. The objective of the system is to assign APs to UEs in a way that maximizes the overall spectral efficiency while controlling interference and preserving fairness. This task, known as scheduling, is formulated here as a contextual decision-making problem. Each decision corresponds to selecting the subset of APs that should serve a given UE under the instantaneous channel and interference conditions.

2.1 Achievable rate

For each UE k, the achievable transmission rate is defined as

$$R_k = log_2(1 + \frac{\sum_{i \in A_k} \beta_{i,k}}{I_{i,k} + \sigma^2}) \tag{16.1}$$

where:

- A_k is the set of APs assigned to UE k,
- $\beta_{i,k}$ is the channel gain between the AP i and the UE k,
- $I_{i,k}$ is the interference from other transmisions,
- σ^2 denotes the noise power.

This formulation captures the dependence of each UE's performance on its serving AP subset, channel conditions, and interference environment.

2.2 Action space

To guide the learning process, the system employs a reward function based on a modified version of the sum-rate:

$$r_w = \sum_{k=1}^{K} R_k + \lambda_1 \text{scaling\_factor} + \lambda_2 \text{penalty}_{\text{action}} \tag{16.2}$$

where additional regularization terms account for scaling and penalize redundant or inefficient AP activations. This formulation encourages the model to maximize spectral efficiency while avoiding suboptimal solutions that could lead to excessive interference or unnecessary complexity.

2.3 Decomposition

Although the global objective is to maximize the system-wide sum-rate, prior work Pereira-Ruisánchez et al. (2025) shows that this can be achieved by maximizing the individual rates R_k. Consequently, the AP assignment problem can be decomposed into independent per-UE subproblems, making it well-suited to contextual bandit formulations. In this setting, each UE is treated as a separate decision unit: given its context vector, the model selects an activation vector, and the corresponding reward is derived from the achieved rate.

This system model provides the foundation for the learning-based approaches developed in the following sections, where Deep Contextual Bandits (DCB) are introduced as the central strategy for dynamic AP assignment in CF-mMIMO systems.

3 Implementation

3.1 Scenario generation

The evaluation is based on simulated CF-mMIMO scenarios generated in two steps:

1. **Physical environment simulation:** APs and UEs are randomly deployed within a predefined area. Distances are computed considering both horizontal and vertical offsets, with APs elevated above UEs. The large-scale channel effects are modeled using standard formulations for pathloss, shadowing, and thermal noise, following 3GPP recommendations 3gp (2019); Goldsmith (2005). The resulting channel gain between AP i and UE k is expressed as
$$\beta_{i,k} = 10^{G_{i,k}/10} \tag{16.3}$$
where $G_{i,k}$ combines pathloss, log-normal shadowing, and noise contributions.
2. **Dataset construction:** From the channel matrix $\beta_{i,k}$, additional features such as interference levels are derived, and a tabular dataset is built. Each sample corresponds to one UE and includes its candidate APs and associated features, forming the context vectors for the learning models.

This setup provides a realistic yet flexible testbed for evaluating resource allocation strategies under varying network conditions.

3.2 Models

Deep Contextual Bandit

The core model treats AP assignment as a contextual bandit problem. Each UE is represented by a context vector containing its channel gains and interference values with respect to candidate APs. Based on this context, a neural network estimates the expected reward of different assignment actions, represented as binary activation vectors. Unlike full reinforcement learning methods, DCB focuses on single-step decisions without temporal correlation, making it efficient and well-suited to fast scheduling in highly dynamic environments.

Additionally, Figure 1 shows how the reward increases steadily during training (a), reflecting the model's improving ability to select good actions. The loss curves of the actor (b) and critic (c) also confirm convergence: while the actor's loss exhibits variability in the early stages due to exploration, the overall trend is decreasing, indicating successful policy adaptation.

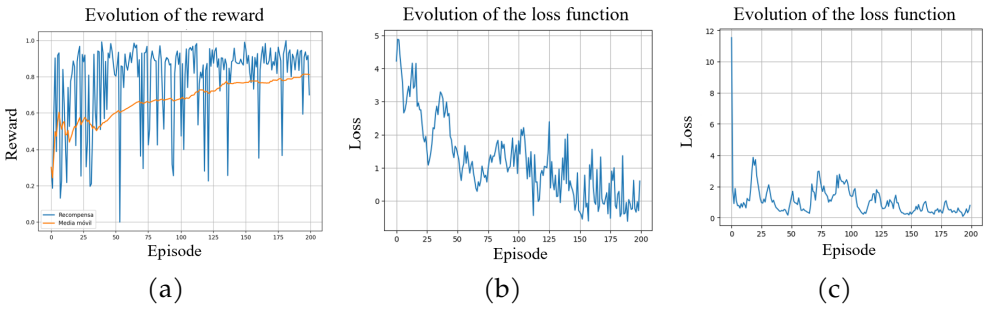

Figure 1: Evolution of the reward and loss functions during training of the DCB model.

Loss-Based Models

Two supervised neural models are included for comparison. Instead of using bandit feedback, these models are trained to directly minimize loss functions derived from the system rate:

- **Inverse-rate loss:** penalizes low rates by minimizing $\frac{1}{R_k}$.
- **Negative-rate loss:** encourages maximization of R_k by minimizing $-R_k$.

These models don't require labeled data from simulated scenarios and provide a complementary perspective, directly aligning learning with rate optimization.

The training results for both variants are shown in Figure 2 The figure illustrates a sharp initial decrease in the loss function followed by stabilization in both cases. This indicates that the models are able to learn efficiently and converge properly.

Fuzzy Clusterinng

An unsupervised fuzzy c-means clustering approach is also implemented. Here, UEs are grouped based on channel conditions, and APs are assigned according to their degree of membership in each cluster. This allows for soft assignments, where multiple APs can serve a UE with weights proportional to their membership strength. A thresholding step ensures practical binary assignments.

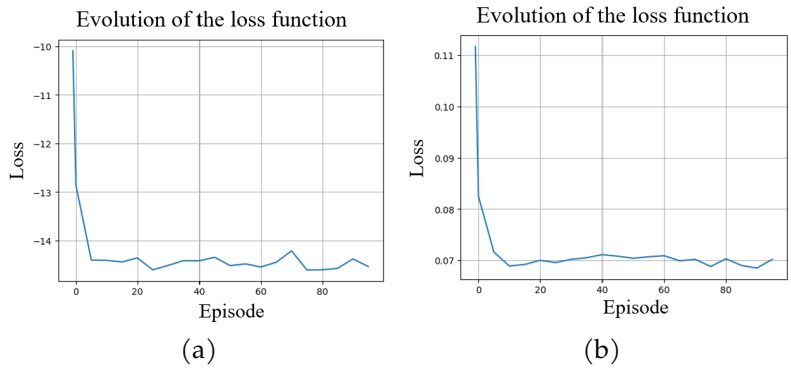

Figure 2: Evolution of the loss function during training of the (a) negative and (b) inverse loss-based models.

4 Results

The models were evaluated across a wide range of simulated CF-mMIMO scenarios, varying the number of APs (L), the number of UEs (K), and the number of candidate APs available per UE (M). In each case, results are averaged over multiple independent realizations to ensure statistical reliability.

4.1 Average Sum-Rate Performance

The following figures show the average system sum-rate achieved by the different approaches across representative scenarios. As expected, the random assignment baseline delivers the lowest performance due to its lack of awareness of channel and interference conditions.

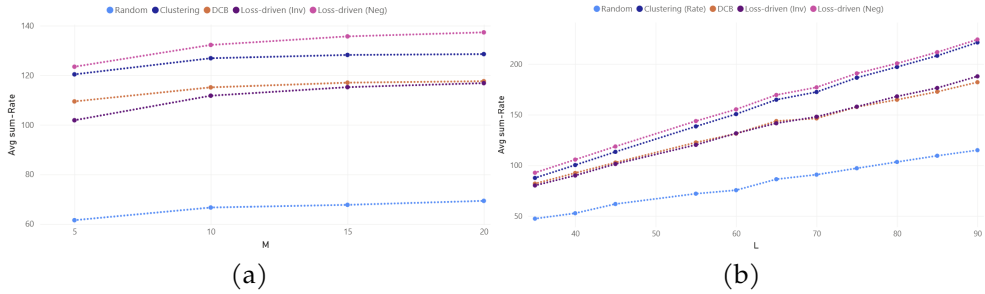

Figure 3: Evaluation results for different values of (a) M and (b) L.

As shown in Figure 3, increasing the number of active APs per UE (M) or the total number of APs (L) expands the set of available connections and degrees of freedom in the system. Here, simpler heuristic or loss-based models can exploit this larger search space more directly, achieving sharper improvements. DCB continues to improve as well, but in a more gradual and conservative way, reflecting its prioritization of stability over aggressive exploitation.

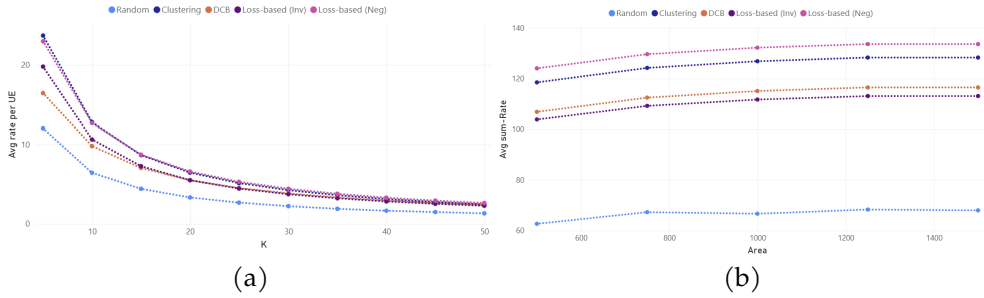

Figure 4: Evaluation results for different values of (a) K and (b) area size.

On the other hand, Figure 4 shows that when the number of users (K) increases, competition for resources becomes stronger and, thanks to its contextual adaptation, DCB handles this congestion more effectively than heuristic models, showing better relative performance. Similarly, when the coverage area expands, the lower density of UEs reduces interference, and DCB's training allows it to adapt flexibly to these conditions, extracting greater benefits than the other approaches.

4.2 Efficiency performance

Figures 5 and 6 show that fuzzy clustering is highly efficient in time and memory, making it suitable for for small to moderate system sizes although it struggles as the number of UEs or APs grows, with memory use rising when L increases. Learning-based models require heavier training but are lightweight at inference, scaling well with many UEs and larger search spaces. In terms of scalability, DCB and loss-based approaches hold an advantage, as their operational cost remains nearly constant after training. DCB consumes slightly more memory than loss-based models due to batch inference, although this increase is negligible from a practical standpoint.

Finaly, the random model, despite its simplicity, is computationally costly because of the vast number of combinations it must evaluate.

Time consumed

Metric	Random	Clustering	DCB	Loss-based (Inv)	Loss-based (Neg)
L	0.16	0.05500	0.12800	0.14000	0.12700
K	0.25	0.01600	0.12200	0.13000	0.11800
M	2.14	0.01800	0.11600	0.11800	0.10900
Total	**2.54**	**0.08900**	**0.36600**	**0.38800**	**0.35400**

Figure 5: Efficiency results in terms of time elapsed.

Memory consumed

Metric	Random	Clustering	DCB	Loss-based (Inv)	Loss-based (Neg)
M	12.40	0.00000	0.00083	0.00006	0.00012
K	1.51	0.00000	0.00010	0.00000	0.00010
L	0.72	1.82700	0.00010	0.00005	0.00000
Total	**14.62**	**1.82700**	**0.00103**	**0.00011**	**0.00022**

Figure 6: Efficiency results in terms of memory consumed.

5 Conclusions

This work addressed the problem of dynamic AP assignment in cell-free massive MIMO systems. We formulated the task as a contextual decision-making problem and proposed a several Machine Learning frameworks to identify AP subsets that maximize spectral efficiency. The main approach, a Deep Contextual Bandits (DCB) model, was compared against loss-based neural models, fuzzy clustering, and a random baseline under a variety of deployment scenarios.

The results demonstrated that DCB achieves the most promising performance. Its ability to make fast, context-aware, per-UE decisions allows it to scale effectively in dense networks. Loss-based models provide competitive performance but are less robust under heterogeneous conditions, with a risk of performance degradation under dynamic or complex environments. Fuzzy clustering, in contrast, only delivers moderate gains over random assignment.

Overall, these findings highlight the potential of contextual bandit methods as lightweight yet powerful tools for scheduling in beyond-5G cell-free networks. Future work will focus on extending the approach to larger-scale deployments, introducing dynamism and temporal correlation to better capture mobility dynamics, broadening the design of reward functions, validating the models on real datasets or standardized simulators, and evaluating robustness against noise in input parameters.

Bibliography

Study on channel model for frequencies from 0.5 to 100 GHz. Technical report, 3GPP, 2019. URL *https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/*.

S. Buzzi, C. D'Andrea, and T. L. Marzetta. User-centric 5g cellular networks: Resource allocation and comparison with the cell-free massive mimo approach. *IEEE*, 2017.

U. Challita, W. Saad, and C. Bettstetter. Machine learning for wireless connectivity and security in the internet of things. *IEEE*, 2019.

A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

F. Liang, C. Zhong, and et al. Deep reinforcement learning for resource allocation in wireless networks. *IEEE*, 2019.

H. Q. Ngo. Cell-free massive mimo. *Encyclopedia of Wireless Networks*, 2020.

H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta. Cell-free massive mimo: Uniformly great service for everyone. *IEEE*, 2017a.

H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta. Cell-free massive mimo versus small cells. *IEEE Transactions on Wireless Communications*, 2017b.

D. Pereira-Ruisánchez, M. Joham, O. Fresnedo, D. Pérez-Adán, L. Castedo, and W. Utschick. Access point assignment for cell-free massive mimo networks using graph neural networks. *IEEE*, 2025.