International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I11P112

Volume 74 Issue 11, 161-173, November 2026
© 2026 Seventh Sense Research Group®

Original Article

Multi-Objective Reinforcement Learning with Graph Neural Networks for Dynamic User-Centric AP Clustering in Cell-Free Ultra-Dense Networks

Paramanand S Patil[1], Satyanarayana.K.Padaganur[2], Santosh Kumbalavati[3]

[1,2,3]*Department of Electronics & Communication Engineering, BLDEA's V.P.Dr.P.G.Halakatti College Of Engineering and Technology Vijayapur, Affiliated to visvesvaraya Technological University Belagavi, Karanatak, India.*

[1]*Coressponding Author: patilparamu@gmail.com*

Received: 17 June 2026 Revised: 10 September 2026 Accepted: 18 September 2026 Published: 25 November 2026

Abstract *- Cell-free networks focused on users require flexibly grouping of access points to enhance the balance between spectral efficiency, energy usage, equity, and operational expenses. This study introduces a reinforcement learning model founded on graph neural networks that simultaneously optimizes these various objectives through a reward structure based on weights. The RL-GNN method is evaluated alongside heuristic, random, Multilayer Perceptron (MLP), Graph Attention Network (GAT), and Weighted Minimum Mean Square-Error (WMMSE) clustering benchmarks under the same channel conditions and patterns of mobility. Comprehensive simulations confirm the capability of RL-GNN to effectively manage and enhance these competing metrics. The system achieves a balanced performance by boosting spectral efficiency, optimizing energy efficiency, ensuring fairness, and maintaining a low reconfiguration rate, demonstrating effective, stable resource allocation in massive MIMO cell-free deployments. Integrating multi-objective reinforcement learning with graph-based models enables context-aware, scalable learning. Overall, user-centric clustering in ultra-dense, cell-free networks shows marked performance gains. The proposed RL GNN in Cell free ultra dense network with user-centric dynamic AP clustering performs pareto optimization of SE, EE, with balanced fairness and reduced reconfiguration compared to baseline clustering methods.*

Keywords *- Reinforcement Learning-based Graphical Neural Network (RL-GNN), Heuristic, Multilayer Perceptron (MLP), Weighted Minimum Mean Square Error (WMMSE), Graph Attention Network (GAT), Ultra Dense Network (UDN).*

1. Introduction

Ultra-dense deployments and cell-free massive MIMO architectures have emerged to address the explosive growth in mobile data traffic and the stringent requirements of latency-sensitive services such as augmented reality, autonomous driving, and real-time health monitoring. Unlike traditional cellular systems that depend on static cell boundaries, the user-centric cell-free paradigm employs a dense set of distributed Access Points (APs) that jointly serve users, enabling seamless handover, finer-grained spatial multiplexing, and significantly improved spectral reuse. The cooperative nature of these APs mitigates inter-cell interference and exploits spatial diversity to boost link reliability and fairness across heterogeneous users and traffic profiles.

However, the benefits come with considerable operational complexity: densely deployed APs generate highly dynamic topologies, large signaling overhead, and variable fronthaul demands. Efficiently forming and adapting AP clusters in real time is therefore critical to unlocking the potential of cell-free ultra-dense networks, optimizing spectral efficiency, reducing energy consumption, and preventing fronthaul bottlenecks-while also ensuring scalability and resilience under mobility and fluctuating traffic patterns. These challenges motivate the development of scalable, context-aware clustering and resource allocation mechanisms that can learn and adapt to environmental conditions with minimal reconfiguration cost. Hence, using RLGNN system can improve the tradeoff between multiple objective parameters.

2. Literature Review

Cell-Free Ultra-Dense Networks (CF-UDNs) are becoming a leading approach for post-5G and 6G wireless systems. Moving away from fixed cell layouts, CF-UDNs rely on coordinated, distributed Access Points (APs) to jointly serve users, which boosts spectral efficiency, delivers more uniform coverage, and cuts down inter-cell interference. These properties make CF-UDNs especially suitable for dense, heterogeneous scenarios [1-4]. In this context, user-centric clustering, where each user is served by a

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

dynamically selected subset of APs chosen according to channel conditions and proximity, has gained wide recognition as an effective means to trade off performance and scalability while also lowering fronthaul load and mitigating pilot contamination [5, 6]. To efficiently manage the complex interactions in CF-UDNs, graph-based learning techniques have proven considerable attention. Wireless networks inherently exhibit graph structures, where nodes (users and APs) are connected through interference and channel relationships. Graph Neural Networks (GNNs) are particularly powerful in capturing such spatial dependencies and have been successfully applied to resource allocation, AP selection, and interference management problems [7-9, 25]. Their ability to generalize across varying network topologies renders them ideal for ultra-dense and dynamic environments. The work done in [16] mainly concentrated on fixed optimization methods within traditional cell-free designs. The study did not take into account user-centered AP clustering when users are moving, learning wireless network structures based on graphs, adaptive clustering using reinforcement learning, or optimization that consider fairness.

In parallel, Reinforcement Learning (RL) has been broadly explored for adaptive resource allocation in wireless systems. Reinforcement Learning (RL) enables agents to learn optimal policies by interrelating with their situation, making it well-suited for dynamic and uncertain network conditions [10, 11]. More recently, Multi-Objective Reinforcement Learning (MORL) has been introduced to address the inherently conflicting objectives in wireless networks, such as maximizing Spectral Efficiency (SE), Energy Efficiency (EE) consumption and ensuring fairness [12-15].Multi-Objective Reinforcement Learning (MORL) frameworks aim to learn Pareto-optimal policies that provide a trade-off among multiple performance metrics.

However, existing research does not primarily focus on specific aspects of these techniques, such as user-centric clustering, GNN-based learning, and Reinforcement Learning (RL) in isolation or through limited partnerships. While some studies use heuristic or optimization-based clustering techniques, others explicitly use deep learning or RL models without explicitly using graph structures [17-19]. Similarly, GNN-based approaches have been applied to AP selection and resource allocation, but often without integrating sequential decision-making capabilities of RL or considering multi-objective optimization [20-24]. Several recent studies have partially combined these techniques, but important gaps remain. For example, DRL-GCN approaches for AP-user association mainly target throughput maximization and omit considerations of fairness and reconfiguration dynamics. Other works employ GNNs to support RL for power control or hardware-aware optimization, yet they stop short of offering a unified framework that jointly handles clustering and multi-objective goals.

Purely GNN-driven methods concentrate on channel modeling or predictive design and do not capture sequential decision making. Spatio-temporal GNNs have been explored for topology optimization, but they generally lack explicit multi-objective reward formulations and mechanisms for dynamic clustering [26-29]. In short, none of these efforts deliver an end-to-end, graph-aware solution that integrates multi-objective optimization with dynamic user-centric clustering. Likewise, recent GNN-based precoder designs [30] optimize physical-layer tasks using learned CSI representations but do not combine reinforcement learning, adaptive clustering, or multi-objective decision making, limiting their use in adaptive CF-UDN deployments. Table 1 lists the identified research gaps motivating proposed work, and the next paragraph summarizes the proposed research.

The organization of the paper is structured as follows: Section 2 glances the Literature review and research gap. Section 3 introduces the system model and formalizes the dynamic user-centric clustering problem, detailing the channel model, network topology, and multiple performance objectives. Section 4 highlights the methodology of multi-objective Graph-aware Neural Networks with Reinforcement Learning (RL-GNN), including algorithm design, state-action space characterization, and reward formulation. Section 5 describes the results and discussion, which include, including simulation set up, including training data generation, baseline comparisons (Heuristics, Random WMMSE, MLP, GAT), and evaluation metrics. provides numerical results illustrating how RL-GNN improves Spectral Efficiency (SE), Energy Efficiency (EE), and Fairness while reducing fronthaul overhead and clustering reconfiguration rates compared to traditional methods. Finally, Section 6 concludes the paper and discusses open challenges and future research directions.

Table 1. Literature review for comparison of existing works with proposed frame work

Ref	Methodology	GNN Used	RL Used	Multi-Objective (SE/EE/Fairness/ Reconfigure)	Dynamic Clustering	Mobility / Temporal Awareness	Key Limitation
[1-6]	Foundational CF / Optimization	✗	✗	✗	✗	✗	No learning, no dynamic clustering
[13, 17, 19]	User-centric clustering / optimization	✗	✗	✗	✓	✗	No learning-based adaptation; static/heuristic clustering

[7, 20, 23-25]	GNN-based AP selection	✓	✗	✗	Partial	✗	No sequential decision-making or multi-objective optimization
[8, 9, 22, 30]	GNN for resource allocation	✓	✗	✗	✗	✗	No clustering + no RL + single-objective focus
[10, 11, 18]	RL-based resource allocation	✗	✓	✗	✗	Partial	No graph modelling → poor scalability in UDN
[12, 14, 15]	Multi-objective RL	✗	✓	✓	✗	Partial	No GNN + no clustering integration
[21]	Multi-agent RL clustering	✗	✓	✗	✓	Partial	No graph structure + lacks multi-objective design.
[27]	DRL + GCN for AP selection	✓	✓	✗	Partial	✗	Focus on throughput only; no fairness/reconfiguration.
[28]	GNN-assisted DRL (power allocation)	✓	✓	Partial	✗	✗	No clustering + limited objectives
[26, 29]	Spatio-temporal GNN	✓	✗	✗	✗	✓	No RL + no reward-driven clustering
[16]	SE-EE optimization	✗	✗	Partial	✗	✗	Analytical optimization only; no learning/adaptation
Proposed Work	RL + GNN for dynamic clustering	✓	✓	✓	✓	✓	Addresses all gaps jointly

2.1. Research Gap

It is observed from the above Table 1 that there is no unified framework that synchronously combines user-centric dynamic clustering, graph neural network-based representation learning, and multi-objective reinforcement learning in cell-free ultra-dense networks. Existing research generally focuses on single-objective aims like throughput, fails to adjust to user movement and fluctuating network situations, or does not clearly model the important trade-offs between Spectral Efficiency (SE), Energy Efficiency (EE), fairness, and reconfiguration overhead.

The integration of context-aware clustering along with mobility, temporal dynamics, and Pareto-optimal decision-making is not commonly found. These gaps leads to the progress of an end-to-end approach that combines reinforcement learning with graph neural networks, aimed at achieving dynamic, scalable, and multi-objective clustering that is centered around user needs in cell-free ultra-dense networks. Summary of the literature review is provided in Table 1. The main contributions of this work are:

- A flexible graph-based approach for representing user-focused clustering in cell-free ultra-dense massive MIMO networks.
- A framework for multi-objective reinforcement learning that simultaneously improves spectral efficiency, energy efficiency, fairness, and reduces reconfiguration cost.
- Incorporating a graph neural network into the reinforcement learning framework to model spatial interference relationships and extract network features that are aware of the topology, enabling adaptive clustering.
- A user-focused algorithm for AP clustering that uses policy-gradient reinforcement learning to enable real-time network adjustments.
- Extensive simulations were conducted using realistic channel and mobility models, with the proposed RL-GNN framework being compared against heuristic methods, as well as MLP and GAT baselines, and WMMSE.

3. System Model and Problem Formulation

3.1. System Model

Let us consider a network with K mobile users, and M distributed APs deployed over a square area with side length D. Positions of users $u_k \in \mathbb{R}^2$and APs $a_m \in \mathbb{R}^2$ advance over discrete-time steps t.

The complex baseband channel from AP M to user K at time t is modelled as $h_{k,m}(t) = \sqrt{\beta_{k,m}(t)}g_{k,m}(t)\beta_{k,m}(t)$ represents large-scale fading (path loss and shadowing) $g_{k,m}(t)$ represents small-scale fading modelled Rayleigh fading i.e$g_{k,m}(t) \sim \mathcal{CN}(0,1)$,each user is dynamically

clustered to a subset of $C_{k(t)} \subseteq \{1, \ldots, M\}$of APs of size up to N_{max} .TDD mode. In downlink transmission, use the same channel model as uplink transmission, It is valid to consider channel reciprocity in TDD systems. The channel-matrix H is used to precode the signal for each user. Figure 1 shows that channel matrix H elements are Complex Gaussian distribution, M array of antennas from access points are linked with K users [1]. Precoding is exploited to map the transmitted signal to the receiver. Figure 2 shows system model for cell-free massive MIMO with M access points and K users [1]. The M access Points transmitted signal is precoded into M×1 vector and then transmit it individually to each terminal of K received terminals over the channel, and detected signal at receiver of each user y=[y1, y2, …yN].

$$y = HTx + n, y \in CK \times 1 \quad (1)$$

$$x = \sqrt{\rho}Pa, x \in CM \times 1 \quad (2)$$

x is transmitted signal for K users in a downlink system with M access points, where P is an M×K linear precoding matrix, a is K×1 transmitted vector before precoding process, and $\sqrt{\rho}$ is the average power of transmitted signal in Figure 1 the P matrix of precoding is related to H is shown.

The generalized precoding block diagram of communication. Figure 2 shows cell free massive MIMO system.

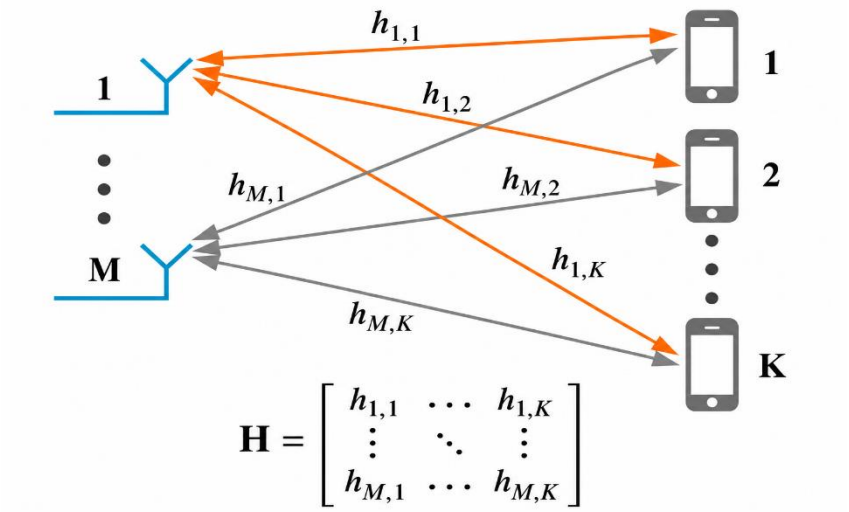

Fig. 1 H spatial matrix for mapping precoding

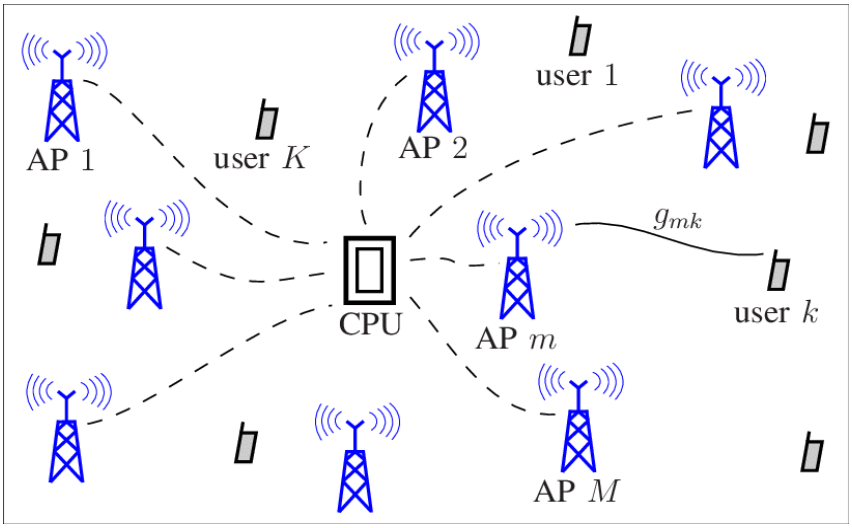

Fig. 2 Cell free massive MIMO model

3.2. Problem Formulation

Formation of user-centric clusters dynamically in real time is essential to maximize spectral and energy efficiency of the network, while minimizing fronthaul and processing overheads. Static cell or cluster boundaries lead to spatial resource wastage and unfair user experiences in dense and

dynamic environments. The current approaches to clustering are based on rules of thumb or supervised learning techniques. These techniques cannot be used to enhance multiple objectives concurrently or adjust well to changes in the network. Dynamic AP clustering requires adaptable algorithms that consider network layout, user movement patterns and interference effects on connections. The challenges include managing the cluster's size relative to the fronthaul load, treating users fairly, decreasing signalling and the costs of changing configurations, and designing a system that can optimize multiple goals simultaneously.

The use of Graph-aware Neural Networks with Reinforcement Learning (RL-GNN) is a technique that involves multi-objective reinforcement learning and addresses these problems. The RL-GNN framework models network graphs for dynamic clustering and learns policies to jointly optimize Energy Efficiency, Spectral Efficiency, Fairness, and reconfiguration overhead. We benchmark obtained method against heuristic, MLP, GAT, and classical WMMSE baselines under realistic mobility and channel conditions. The contributions include the design of the RL-GNN architecture tailored to cell-free UDNs, multi-objective reward function formulation, and comprehensive performance evaluation validating real-time adaptive clustering feasibility. The GNN architecture captures spatial and topological dependencies across access points and users, allowing the RL agent to jointly optimize spectral efficiency, energy efficiency, and fairness, while minimizing clustering reconfiguration rate and fronthaul load. Unlike conventional optimization models or learning-based clustering, the proposed RL-GNN operates in real time with minimal centralized coordination, supporting user mobility and dynamic topology changes efficiently. The utility function developed to maximize is noted in bellow Equation (3), and the related equations to quantify the multi-objective constraints to scale the performance parameters over multiple iterations.

$$\mathrm{U(t)} = \alpha.\,\mathrm{SE(t)} + \beta.\,\mathrm{EE(t)} + \gamma.\,\mathrm{F}(\{\gamma_{\mathrm{k}}(\mathrm{t})\}) - \delta.\,\mathrm{C}_{\mathrm{reconfigure(t)}} \quad (3)$$

Where α, β, γ, are weighting or scaling factors in multi-objective optimization, giving the importance to each variable SE, Energy, Fairness and reconfiguration rate contributing for reward calculation.

α, β, γ, δ ≥ 0, $\bar{R}(t)$ is spectral efficiency, and F is Jains fairness index of user rates $\gamma_{\mathrm{k}}(t)$, E(t) is total power spent by active APs, $C_{reconfigure(t)}$ quantifies the cost of cluster changes from t-1 to further these performance points are noted in below equations.

$$SE(t) = B\log_2\left(1 + \mathrm{SINR}_k(t)\right) \quad (4)$$

Where $\mathrm{SINR}_k(t)$ represents the Signal to Interference plus Noise Ratio (SINR) for user K at time t .

$$\mathrm{SINR_k}(t) = \frac{\left|\sum_{m\in\mathcal{C}_k(t)} h_{k,m}(t)\, w_{k,m}(t)\right|^2}{\sum_{j\neq k}\left|\sum_{m\in\mathcal{C}_j(t)} h_{k,m}(t)\, w_{j,m}(t)\right|^2 + \sigma^2} \quad (5)$$

$w_{k,m}(t)$ denotes precoding weights, $h_{k,m}(t)$ denotes channel coefficient from AP M, → user K, σ^2 denotes noise power.

$$F\left(\{\gamma_{k(t)}\}\right) = \frac{\left(\sum_{k=1}^{K}\gamma_k(t)\right)^2}{K\sum_{k=1}^{K}\left(\gamma_k(t)\right)^2} \quad (6)$$

$F\left(\{\gamma_{k(t)}\}\right)$ is Jains fairness index for user rates$\gamma_{\mathrm{k}}(t)$.

Power spent by active Access points over M Access points

$$EE(t) = \sum_{m=1}^{M}[a_m(t)P_{\text{active}} + \left(1 - a_m(t)\right)P_{\text{sleep}}] \quad (7)$$

$a_{m(t)}$ =1 *AP m serves at least one user at time t.*

Otherwise,

$$C_{reconfigure(t)} = \frac{1}{M\,K}\left(\Sigma_{m=1}^{M}\Sigma_{k=1}^{K}\left|\,C_{(m,k)(t)} - C_{(m,k)(t-1)}\right|\right) \quad (8)$$

The multi-objective utility function maximization $U(t)$ subjected to constraints $|\bar{R}(t)| \leq$ Nmax for all k ∀k (Cluster size constraint)

Fronthaul (t) ≤ Fmax (Front Haul Budget)

Pm(t) ≤ Pmax for ∀m (Power Constraint)

Consistency: $|\bar{\mathrm{R}}(\mathrm{t})|$ Coherent across overlaps.

RL-GNN-based Multi-Objective optimization with expected discounted cumulative reward is used to solve dynamic optimization, using a reinforcement learning framework where the environment state $[s_t]$, encodes a graph [Gt] containing APs, users, and channel features. The GNN policy $[\pi_\theta(a_t|s_t)]$ outputs cluster selection probabilities, where actions $[a_t]$ correspond to AP user associations. The policy parameters [θ] are updated via policy-gradient methods to maximize the expected discounted reward, where the reward rt follows the multi-objective utility Ut GNN layers extract graph-based network features, enabling spatial-temporal reasoning across users and APs.

$$J\,(\theta) = \max_{\pi\theta}\,\mathbb{E}_{\pi_\theta}\,[\textstyle\sum_{t=0}^{T}(\gamma^t r_t)] \quad (9)$$

$\gamma \in [0,1)$ The discount factor determining the relative importance of future rewards. rt is reward maximizing the utility [Ut].

4. Methodology

A cell-free ultra-dense massive MIMO network consisting of geographically distributed Access Points (APs) serving multiple mobile users is considered. The objective is to dynamically determine user-centric AP clusters that jointly maximize Spectral Efficiency (SE) and Energy Efficiency (EE) while maintaining fairness and reducing cluster reconfiguration overhead in mobility-aware wireless environments.

Initially, APs and users are randomly deployed within the simulation region, and user mobility is modelled over discrete time intervals to capture temporal variations in network topology. At each time instant, the wireless channel is generated by incorporating small-scale fading, large-scale path loss, shadowing, and spatial correlation effects. Based on the resulting channel realizations, a graph-based representation of the network is constructed, where nodes correspond to APs and users, and edges characterize wireless link conditions.

Node and edge features, including channel gain, AP-user distance, interference level, and connectivity statistics, are extracted and aggregated to form the Reinforcement Learning (RL) state representation. To adapt to user mobility and channel dynamics, the proposed framework employs a Proximal Policy Optimization (PPO)-based actor-critic architecture for dynamic clustering. The actor network generates user-centric AP clustering decisions according to the current network state, whereas the critic network evaluates the long-term utility of the selected actions to facilitate policy improvement. Both actor and critic networks are implemented as fully connected multilayer perceptrons with ReLU activation functions.

PPO ensures stable policy updates by constraining excessive deviations from previously learned policies during training. Based on the selected action, a user-centric clustering matrix is constructed under predefined cluster-size constraints, ensuring that each user is associated with a limited number of serving APs. Only APs belonging to active clusters participate in data transmission, thereby reducing unnecessary energy consumption. Subsequently, Minimum Mean-Square Error (MMSE) precoders are computed for all active AP-user links and incorporated into the physical-layer transmission model. The resulting signal model is used to evaluate received signal power, inter-user interference, Signal-to-Interference-Plus-Noise Ratio (SINR), per-user spectral efficiency, and aggregate network throughput.

The energy efficiency evaluation considers not only transmit power but also circuit power, Digital Signal Processing (DSP) power, and cluster reconfiguration energy. Consequently, the framework provides a comprehensive assessment of dynamic clustering performance while accounting for practical physical-layer and implementation-related energy costs. The RL agent is trained using a multi-objective reward function designed to simultaneously maximize spectral efficiency, energy efficiency, user fairness, and clustering stability.

The reward function is expressed as,

$$R(t) = \alpha SE(t) + \beta EE(t) + \gamma F(t) - \delta C_{reconfig(t)}$$

The wireless network at time t is represented as a bipartite graph$\left(G_t = (V, E_t)\right)$, where the node set $V = V_u \cup V_a$consists of user nodes and AP nodes. An edge represents $(a_m, u_k) \in E_t$ a wireless link characterized by the channel coefficient $h_{m,k}(t)$.

The graph-aware feature set includes channel gains, large-scale fading, interference levels, user locations, and AP connectivity information, all of which are used to build the RL state for dynamic AP clustering. During each training episode, the PPO agent interacts with the wireless environment by observing graph-based state information, choosing clustering actions, receiving PHY-layer rewards, and updating the actor-critic parameters.

Through this iterative process, the framework learns a clustering policy that adapts to user mobility, interference changes, and evolving network conditions. The PPO policy generates clustering actions that determine AP-user association decisions. The policy parameters are restructured using policy-gradient optimization to maximize the expected discounted cumulative reward.

$$max_{\pi\theta} E_{\pi\theta}[\Sigma_{t=0}^{T} \gamma^t R(t)] \quad (10)$$

Where,

$\pi\theta$ = PPO policy parameterized by θ,

γ = discount factor,

R(t) = multi-objective reward at time t.

The overall RL-GNN framework operates iteratively through:

- Channel generation,
- Graph feature extraction,
- PPO-based clustering,
- MMSE precoding,
- PHY-aware SINR evaluation,
- Reward computation,
- Policy update.

This enables adaptive and scalable AP clustering under dynamic wireless conditions.

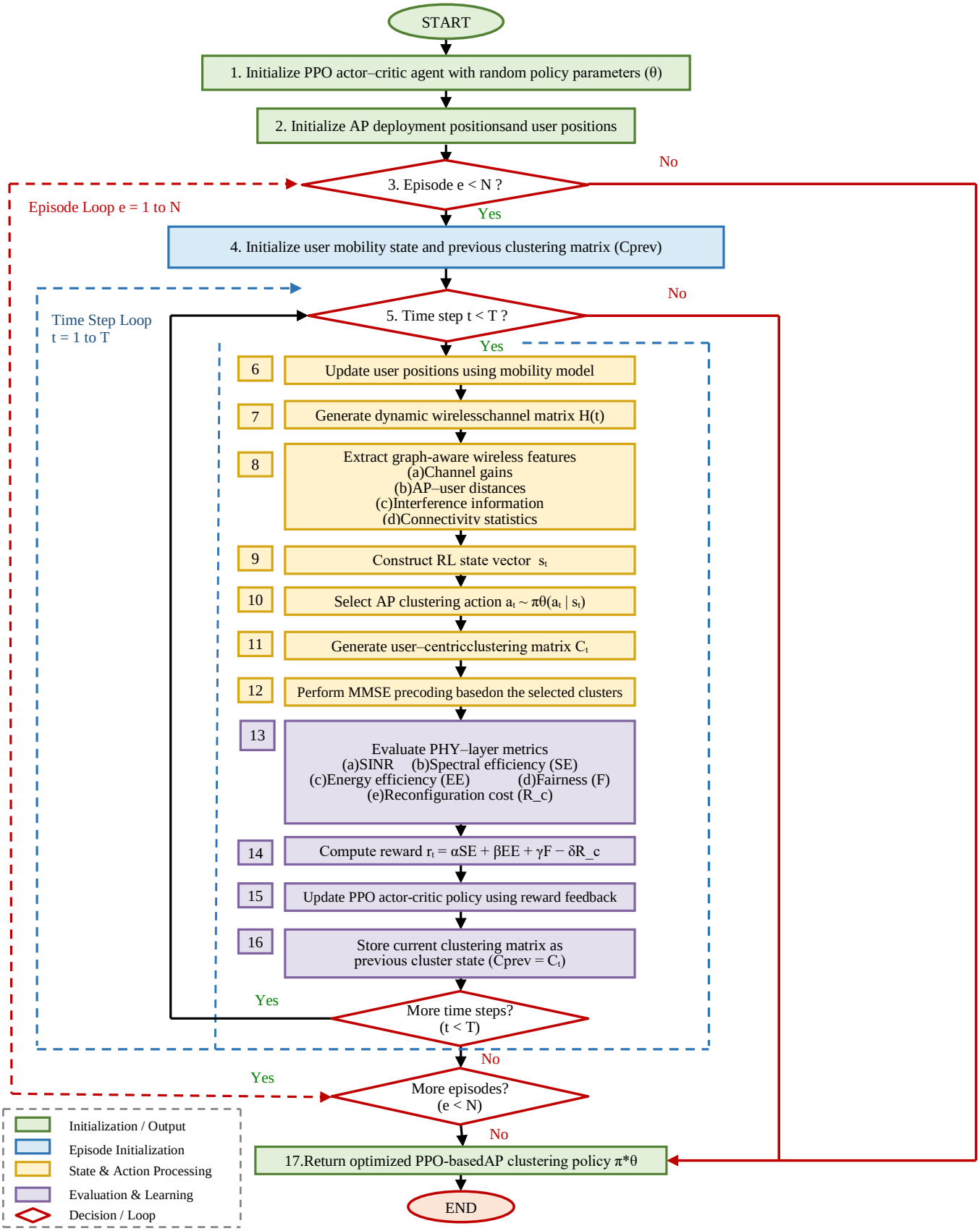

Fig. 3 Algorithm 1 flow chart

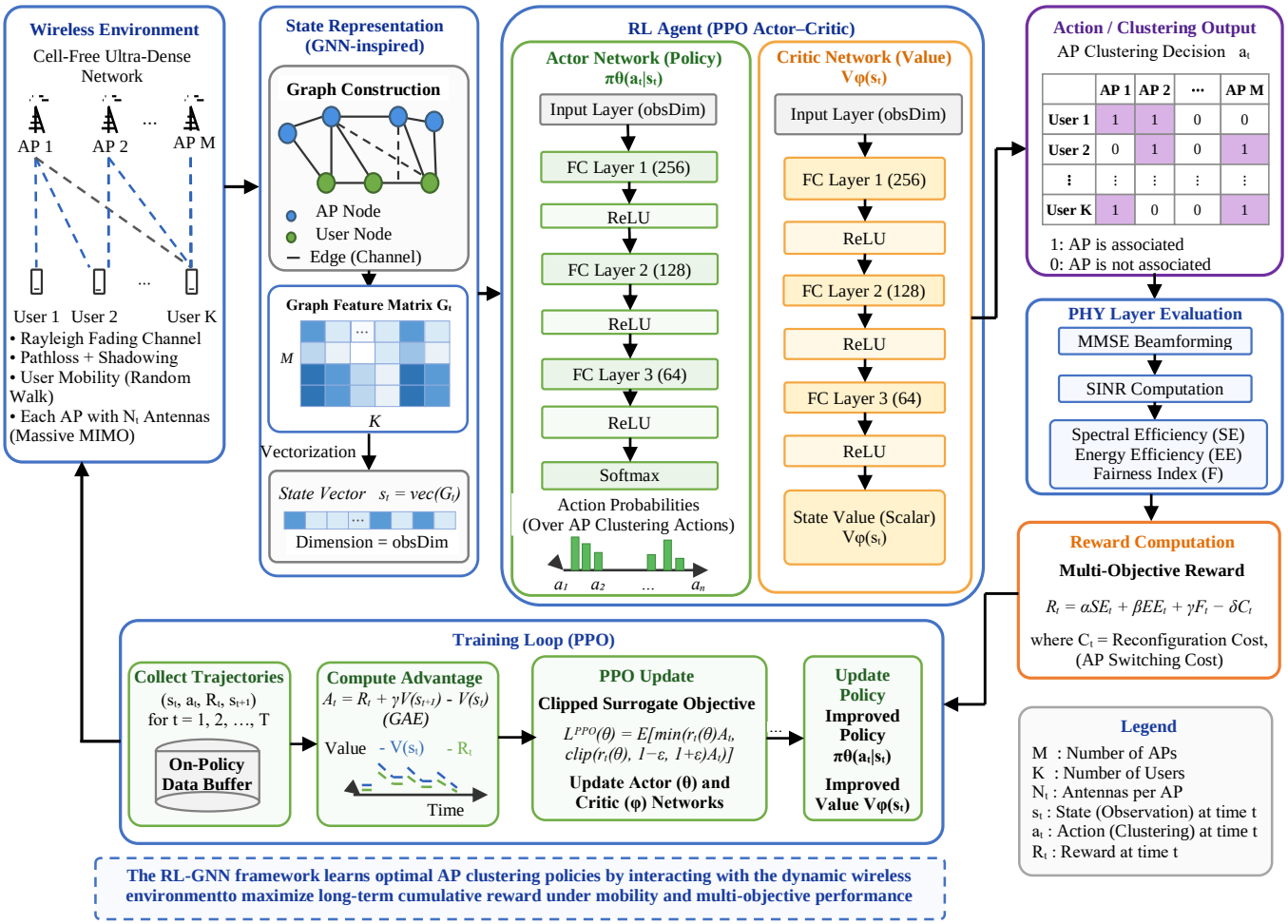

Fig. 4 Architecture of the RL-GNN for dynamic user centric a clustering in Cell Free-Ultra Dense Networks (CF-UDNs)

Algorithm 1. PPO-based RL Framework for Dynamic User-Centric AP Clustering

Input:

Number of APs (M), users (K), antennas per AP(N_t) maximum episodes (N), maximum timesteps (T), PPO policy parameters θ.

Output:

Optimized PPO-based AP clustering policy π_{θ^*}

Step 1: Initialize PPO actor-critic agent with random policy parameters (\theta).

Step 2: Initialize AP deployment positions and user positions.

Step 3: for episode (e=1) to (N) do

Step 4: Initialize user mobility state and previous clustering matrix.

Step 5: for timestep (t=1) to (T) do

Step 6: Update user positions using the mobility model.

Step 7: Generate dynamic wireless channel matrix (H(t)).

Step 8: Extract graph-aware wireless features, including:

a) channel gains,
b) AP-user distances,
c) interference information,
d) connectivity statistics.

Step 9: Construct RL state vector s_t.

Step 10: Select AP clustering action

Step 11: Generate user-centric clustering matrix C_t.

Step 12: Perform MMSE precoding based on the selected clusters.

Step 13: Evaluate PHY-layer metrics:

a) SINR,
b) spectral efficiency,
c) energy efficiency,
d) fairness,
e) reconfiguration cost.

Step 14: Compute reward function r(t) = α.SE(t) + β.EE(t) + γ.F(t) - δ.Reconfig(t)

Step 15: Update PPO actor-critic policy using reward feedback.

Step 16: Store current clustering matrix as previous cluster state.

Step 17: end for

Step 18: end for

Step 19: Return Optimized PPO-Based AP Clustering Policy $\pi\theta^*$

Step 20: End

The complete implementation flow of the proposed RL-based dynamic AP clustering framework is summarized in Algorithm 1. Flowchart of algorithm is shown in Figure 3.

The PPO-based actor and critic networks consist of three fully connected hidden layers with 256, 128, and 64 neurons, each using ReLU activation. Both networks accept the observation vector at the input layer and produce an output layer, where the actor uses SoftMax to generate action probabilities and the critic outputs a scalar state-value estimate.

The proposed RL-GNN framework integrates a Proximal Policy Optimization (PPO)-based actor-critic architecture to perform user-centric access point clustering in cell-free ultra-dense networks. The state representation is constructed from network observations, including Channel State Information (CSI), interference measurements, user mobility characteristics and current cluster configurations. Within the learning architecture, intermediate network layers extract latent representations that capture channel dynamics, graph-structured relationships among users and access points, and the trade-offs associated with multiple optimization objectives. These learned representations are subsequently utilized by the actor network to determine AP clustering decisions that adapt to variations in user location and network conditions.

The critic network estimates. The critic network estimates the expected cumulative reward corresponding to the selected clustering action and provides feedback for policy refinement during training. Through this evaluation process, the learning agent seeks to improve spectral efficiency, energy efficiency, user fairness, and clustering stability simultaneously. To ensure stable learning, PPO updates the policy using the clipped surrogate objective function defined in (11), thereby limiting excessive policy variations and improving convergence during dynamic AP clustering. This clipping limits large changes between successive policies and helps prevent unstable AP reconfiguration under rapidly changing user mobility and channel conditions, enabling smooth convergence and robust learning in highly dynamic cell-free ultra-dense network environment.

$$L^{CLIP}(\theta) = \hat{E}_t\left[\min\left(r_t(\theta)\widehat{A_t}, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\widehat{A_t}\right)\right] \quad (11)$$

This Equation (11) controls how much the policy can change during training.

4.1. Implementation Details

The simulation environment is implemented in MATLAB using the Reinforcement Learning Toolbox, simulating an urban square area of size 200m × 200m. K users and M distributed Access Points (APs) are randomly initialized within the area. User mobility is simulated as a Gaussian random walk with bounded velocity to mimic realistic movement. Propagation channels between APs and users follow a stochastic Rayleigh fading-channel model with large-scale path loss dependent on distance as per the log-distance path loss model.

Channel matrices $H(t) \in C^{K\times M}$ are updated each time step t,The Environment state comprises the instantaneous channel matrix combined with user and AP positional data, represented as node and edge features in a graph. Actions correspond to user-centric clusters formed through the dynamic selection of AP subsets limited by a cluster size $N_{max} = 5$ environment's step method updates user positions, channel realizations, and evaluates the multi-objective reward function, aggregating Spectral Efficiency (SE), Energy Efficiency, fairness, energy consumption, and reconfiguration cost. the cell free wireless network simulation parameters in the MATLAB environment are mentioned in Table 2.

Table 2. RL-GNN framework configuration, PPO hyperparameters, and simulation settings

Component	Description	Configuration
Number of APs	Access Points	M=20
Number of Users	Active Users	K=40
Antennas per AP	Massive MIMO Configuration	Nt=4
RL Algorithm	Reinforcement Learning Method	Proximal Policy Optimization (PPO)
RL Type	Learning Paradigm	Policy Gradient Actor-Critic
Environment	Wireless System	Cell-Free Ultra-Dense Network
State Representation	Input Features	Graph-aware AP-user features
State Dimension	Observation Size	(M × K)
Action Space	AP Clustering Decisions	Dynamic AP Selection
Actor Network	Policy Network	Fully Connected Neural Network
Critic Network	Value Network	Fully Connected Neural Network
Input Layer	Observation Input	ImageInputLayer([obsDim 1 1])

Hidden Layer 1	Feature Extraction	256 Neurons + ReLU
Hidden Layer 2	Feature Extraction	128 Neurons + ReLU
Hidden Layer 3	Feature Extraction	64 Neurons + ReLU
Actor Output Layer	Action Probability	Fully Connected Layer + Softmax
Critic Output Layer	State Value Estimation	Fully Connected Layer (1 Neuron)
Activation Function	Nonlinearity	ReLU
Policy Output	Action Selection	Stochastic Softmax Policy
Optimizer	PPO Optimizer	Adam
Learning Rate	PPO Learning Step	0.001
Discount Factor	Future Reward Weight	0.99
Experience Horizon	PPO Horizon Length	128
Clipping Factor	PPO Stability Parameter	0.2
Entropy Weight	Exploration Control	0.01
Maximum Episodes	RL Training Episodes	50-200
Reward Function	Multi-Objective Optimization	SE + EE + Fairness − Reconfiguration Cost
Beamforming Method	PHY Layer Precoding	MMSE
Mobility Model	User Mobility	Random Walk / Dynamic Mobility
Channel Model	Wireless Channel	Rayleigh Fading with Pathloss
AP Clustering	User-Centric Clustering	Dynamic
Performance Metrics	Evaluation Metrics	SE, EE, Fairness, Reconfiguration Cost
Cluster Size	Maximum Serving APs per User	Ck=5

4.2. Complexity, Execution-Time, and Modularity Analysis

The computational complexity of the proposed PPO-driven dynamic AP clustering framework is dominated by wireless channel synthesis, MMSE precoder computation, PHY-layer metric evaluation, and PPO policy updates. In a network with M APs, K users, and N_t antennas per AP, MMSE precoding, because of matrix inversions, constitutes the primary cost, approximately O(KNt)3. The PPO update cost is governed by the actor-critic neural networks' forward and backward propagation.

Per time step, the total execution complexity can be approximated as : O(MKNt+(KNt)3+Np), where Np denotes the computational cost of PPO neural-network updates. The framework was implemented in MATLAB R2020a and run on a standard workstation. PPO training showed stable convergence with low reward variance; depending on user mobility and network density, most runs converged in about 150-250 episodes. Overall runtime is primarily influenced by the numbers of APs and users, the count of Monte Carlo trials, and the number of PPO training episodes.

5. Results and Discussion

The results illustrate the proposed dynamic clustering method. based on RLGNN performs the best compared to all the other methods that were looked at. When it comes to Spectral Efficiency (SE), RLGNN reaches the highest value of 83.29 bps/Hz, which is better than traditional methods like Heuristic (79.522), WMMSE (79.585), GAT (44.896), and MLP (32.272). This improvement shows that the RL-GNN framework can effectively enhance how access points connect with users and how resources are distributed. In the same way, the suggested RLGNN reaches the best Energy Efficiency (EE) of 5.49 bits per Joule, which is a little bit higher than WMMSE's 5.4777 and much better than the MLP and GAT methods.

Table 3. Final results

Method	SE_ avg	EE_avg	Fairness_avg	Reconfig_avg
Heuristic	79.522	5.2357	0.42191	199.8
MLP	32.272	2.6973	0.18866	131.8
GAT	44.896	2.9612	0.27129	198.9
WMMSE	79.585	5.4777	0.40686	185.4
RLGNN(Proposed)	83.29	5.49	0.43908	190

Table 3 shows the final comparative results obtained for proposed system with baseline methods with respect to performance parameters dynamic clustering APs. This shows that the smart learning system based on graphs can improve Spectral Efficiency (SE) while keeping power usage low. For fairness performance, RLGNN records the highest Jain's fairness index of 0.43908. This shows that it provides a more even distribution of resources among users when compared to other methods. Even though the heuristic method shows good performance in SE, it is not as fair as the RLGNN approach we suggested. When looking at the costs of changing the setup, the MLP method has the lowest average reconfiguration

score at 131.8. However, this minimum cost means that there is a big drop in both SE and EE performance. The suggested RLGNN strikes a nice balance between improving performance and the cost of reconfiguring, leading to better overall network optimization than both traditional methods and those based on learning.

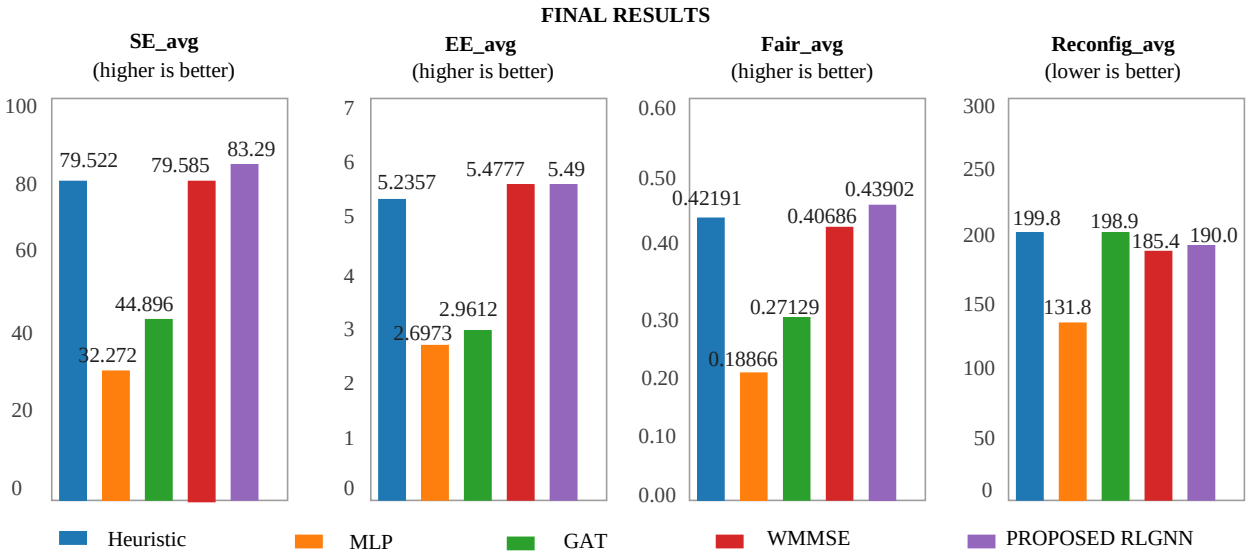

Fig. 5 Comparative performance parameters in Cell Free-Ultra Dense Networks (CF-UDNs)

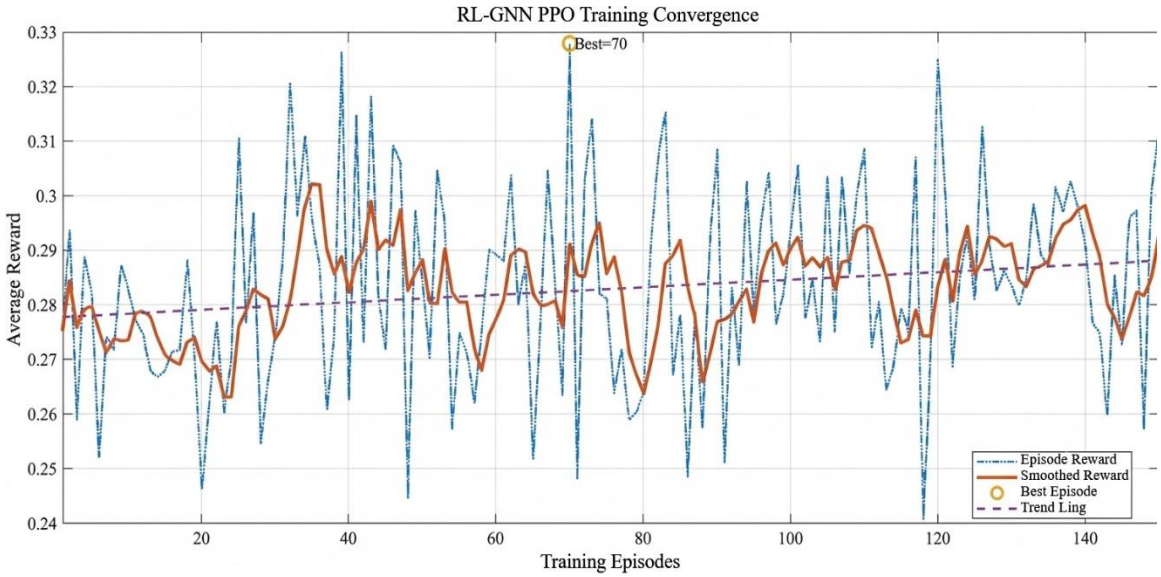

Fig. 6 RL-GNN PPO training convergence

The comparative trade off of performance parameters are shown in Figure 5. Figure 6 illustrates the training convergence behaviour of the proposed RL-GNN PPO framework. The episodic reward exhibits noticeable fluctuations due to the inherent randomness of mobility and wireless channel dynamics; however, the smoothed reward shows a consistent upward trend throughout training, indicating progressive policy improvement. The best performance is obtained around episode 70, after which the reward stabilizes within a narrow interval, suggesting near-convergence of the learned clustering policy. This behaviour confirms that the proposed RL-GNN framework can effectively learn adaptive AP clustering decisions under dynamic network conditions while maintaining stable PPO training.

6. Conclusion

We propose a PPO-based reinforcement learning framework for dynamic, user-centric AP clustering in cell-free ultra-dense massive MIMO systems. The method jointly balances spectral efficiency, energy efficiency, fairness, and cluster reconfiguration cost in mobility-driven, time-varying

wireless environments. Graph-informed state features such as channel gains, interference metrics, and AP-user connectivity statistics-are embedded into the RL observation to enhance adaptive clustering decisions.

Simulation studies show the RL framework outperforms heuristic methods, MLP-based and GAT-based baselines, and conventional WMMSE in both spectral and energy efficiency while preserving comparable fairness. PPO training demonstrates stable convergence with low reward variance. The modular MATLAB implementation facilitates the integration of different mobility models, precoding techniques, and optimization objectives. Although the approach incurs extra computational overhead from PPO training and MMSE precoding, it delivers superior adaptability and scalable operation suitable for future intelligent wireless systems. Potential extensions include:

- Distributed multi-agent RL,
- GNN-assisted feature learning,
- Federated clustering approaches,
- Modelling hardware impairments,
- Latency- and URLLC-aware optimization,
- Real-time edge intelligence deployment,
- Energy-harvesting and green-communication models.

The framework is amenable to evolution toward beyond-5G/6G cell-free architectures for autonomous radio-resource management in very large-scale, ultra-dense networks.

Conflicts of Interest

"The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper."

Funding Statement

Authors state that no funding is received by any of the agency or organisations.

Acknowledgement

We are thankful to principal BLDEA's V.P.Dr.P.G. Halkatti college of engineering by encouraging and supporting for this research initiative, and also thank all the management personnel of BLDEACET Vijayapur.

References

[1] Qingjiang Shi et al., "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331-4340, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[2] Hien Quoc Ngo et al., "Cell-Free Massive MIMO Versus Small Cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834-1850, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[3] Elina Nayebi et al., "Performance of Cell-Free Massive MIMO Systems with MMSE and LS Channel Estimation," *2016 50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, pp. 203-207, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[4] Giovanni Interdonato et al., "Ubiquitous Cell-Free Massive MIMO Communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1-13, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Emil Björnson, and Luca 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. [CrossRef] [Google Scholar] [Publisher Link]

[6] Emil Björnson, and Luca Sanguinetti, "Scalable Cell-Free Massive MIMO Systems," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247-4261, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Franco Scarselli et al., "The Graph Neural Network Model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61-80, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mark Eisen, and Alejandro Ribeiro, "Optimal Wireless Resource Allocation with Graph Neural Networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 538-553, 2020. [CrossRef] [Publisher Link]

[9] Shiwen He et al., "An Overview on the Application of Graph Neural Networks in Wireless Networks," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 2547-2565, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Qingqing Wu, Yong Zeng, and Rui Zhang, "Joint Trajectory and Communication Design for Multi-UAV Enabled Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 2109-2121, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Adam Girycki et al., "Learning-based Precoding-Aware Radio Resource Scheduling for Cell-Free mMIMO Networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 5, pp. 4876-4888, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Amal Feriani, and Ekram Hossain, "Single and Multi-Agent Deep Reinforcement Learning for AI-Enabled Wireless Networks: A Tutorial," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 2, pp. 1226-1252, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Stefano Buzzi, and Carmen D' Andrea, "Cell-Free Massive MIMO: User-Centric Approach with Distributed Optimization," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 706-709, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[14] Youngwoo Oh, Arif Ullah, and Wooyeol Choi, "Multi-Objective Reinforcement Learning for Power Allocation in Massive MIMO Networks: A Solution to Spectral and Energy Trade-Offs," *IEEE Access*, vol. 12, pp. 1172-1188, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Guorong Zhou et al., "Joint Multi-Objective Optimization for Radio Access Network Slicing using Multi-Agent Deep Reinforcement Learning," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 11828-11843, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Han Wang et al., "Joint Optimization of Spectral Efficiency and Energy Efficiency with Low-Precision ADCs in Cell-Free Massive MIMO Systems," *Science China Information Sciences*, vol. 65, no. 5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yanpeng Dai et al., "A Survey of Graph-based Resource Management in Wireless Networks-Part II: Learning Approaches," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 4, pp. 2101-2122, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[18] Charmae Franchesca Mendoza, "*Deep Reinforcement Learning for Cell-Free Massive MIMO Network Optimization*," Thesis Dissertation, Institute of Telecommunications, Vienna University of Technology, Vienna, Austria, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Shuaifei Chen et al., "A Survey on User-Centric Cell-Free Massive MIMO Systems," *Digital Communications and Networks*, vol. 8, no. 5, pp. 695-719, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] S. Salehi et al., "Machine Learning-based AP Selection in User-Centric Cell-Free Multiple-Antenna Networks," *arxiv*, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Bitan Banerjee et al., "Access Point Clustering in Cell-Free Massive MIMO using Conventional and Federated Multi-Agent Reinforcement Learning," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 1, pp. 107-123, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Yanpeng Dai et al., "A Survey of Graph-based Resource Management in Wireless Networks—Part II: Learning Approaches," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 4, pp. 2101-2122, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Yu Zhao et al., "A Graph Convolutional Network Approach for AP Selection in User-Centric Cell-Free Massive MIMO," *2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE)*, Guangzhou, China, pp. 335-339, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Vismika Ranasinghe, Nandana Rajatheva, and Matti Latva-Aho, "Graph Neural Network based Access Point Selection for Cell-Free Massive MIMO Systems," *2021 IEEE Global Communications Conference (GLOBECOM)*, Madrid, Spain, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Sabarish Krishna Moorthy, and Jithin Jagannath, "Survey of Graph Neural Network for Internet of Things and NextG Networks," *IEEE Communications Surveys and Tutorials*, vol. 28, pp. 1803-1844, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Diego Hortelano et al., "A Comprehensive Survey on Reinforcement-Learning-based Computation Offloading Techniques in Edge Computing Systems," *Journal of Network and Computer Applications*, vol. 216, pp. 1-35, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Mingjun Du et al., "Deep Reinforcement Learning-Based Access Point Selection for Cell-Free Massive MIMO with Graph Convolutional Networks," *2023 IEEE 17th International Conference on Future Communication Networks (FCN)*, Queenstown, New Zealand, pp. 1-5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Meng Zhou et al., "GNN-Assisted Deep Reinforcement Learning for Cell-Free Massive MIMO Systems with Nonlinear Power Amplifiers and Low-Resolution ADCs," *IEEE Internet of Things Journal*, vol. 12, no. 16, pp. 33041-33055, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[29] Jing Jiang et al., "Spatio-Temporal GNN-based Cell-Free Massive MIMO Network with Maximal Benefit-Cost Ratio," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 7, pp. 11490-11494, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[30] Tianzheng Miao et al., "GNN-based Precoder Design and Fine-tuning for Cell-free Massive MIMO with Real-world CSI," *arXiv*, pp. 1-6, 2025. [CrossRef] [Google Scholar] [Publisher Link]