International Journal of Basic and Applied Sciences, 14 (4) (2025) 475-487

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/2z0jsb23
Research paper

Enhancing UAV Network Efficiency: A Multi-Agent DRL Approach for Joint User Association And Trajectory Optimization

Vijaya Babu Burra [1] *, Vullam Naga Gopi Raju [2], Banda Snv Ramana Murthy [3], Rohini Rajesh Swami Devnikar [4], Sarala Patchala [5], Suneetha Jalli [5], Vasanthi Yarra [6], Vinnakota Naga Venkata Swathi [7], Desamala Prabhakar Rao [8], Yarrapragada K. S. S. Rao [3]

[1] Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, GUNTUR (DT)
[2] PROFESSOR, Department of Computer Science and Engineering, Chalapathi Institute of Engineering and Technology, Guntur
[3] Assistant Professor, Department of CSE-AIML, ADITYA UNIVERSITY, SURAMPALEM, A. P.
[4] Assistant Professor, G H Raisoni College of Engineering and Management, Pune
[5] Associate Professor, Department of ECE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh Pradesh, India, Andhra Pradesh, India
[6] Assistant Professor, Department of Computer Science and Engineering (AI&ML), R.V. R & J.C. College of Engineering, Andhra Pradesh, India
[7]Assistant Professor, Department of Computer Science and Engineering (AI&ML), R.V. R & J.C College of Engineering, Andhra Pradesh, India
[8] Department f ECE, Chalapathi Institute of Technology, Mothadaka, Guntur, Andhra Pradesh, India
[3] Department of Mechanical Engineering, Aditya University, Surampalem, Andhra Pradesh, India
**Corresponding author E-mail: vijay_gemini@kluniversity.in*

Received: June 18, 2025, Accepted: August 1, 2025, Published: August 16, 2025

Abstract

In multi-UAV networks, user equipment (UE) needs to connect to UAVs for both uplink (UL) and downlink (DL). Traditionally, UL and DL associations are coupled means a UE connects to the same UAV for both. However, this approach is inefficient due to the mobility of UAVs and network heterogeneity. Full-duplex (FD) communication in UAV networks complicates this problem. This paper introduces a novel decoupled UL-DL association (DUDe) framework. This allows each UE to associate with different UAVs for UL and DL transmissions. Such decoupling improves flexibility and enhances communication performance. However, UE association depends on UAV trajectories, making the prob-lem more complex. However, the dependence of UE association on UAV trajectories increases the complexity. This work formulates a joint optimization problem to maximize the total sum-rate of UEs in both UL and DL. To handle this uncertainty, a robust Partially Observable Markov Decision Process (POMDP) model is used. This helps model the uncertain environment where UAVs do not have complete information about the system state. A Multi-Agent Deep Reinforcement Learning (MADRL) ap-approach is proposed to solve this problem. Each UAV selects its policy in a decentralized manner. The training process is improved using a modified Proximal Policy Optimization (PPO) algorithm. It uses deep reinforcement learning to efficiently manage UAV mobility and UE connectivity. The findings suggest that DUDe-based associations outperform traditional coupled associations. It leads to better spectral efficiency and higher network throughput. The proposed framework and algorithms are validated through simulations and real-world scenarios.

Keywords: *UAV; Network; Efficiency; UL; DL; POMDP.*

1. Introduction

The increased demand for high-speed and reliable wireless communication has led to significant advancements in network technologies [1]. Traditional terrestrial networks rely on fixed base stations (BSs) to provide coverage and connectivity. However, deploying these BSs incurs high costs and lacks flexibility [2]. In emergencies, including natural disasters or temporary events, deploying terrestrial BSs becomes impractical. To overcome these limitations, Unmanned Aerial Vehicles (UAVs) have emerged as a promising solution [3]. UAVs are rapidly deployed, providing wireless communication services with high mobility and adaptability. UAV-based networks offer several advantages [4]. Firstly, UAVs are positioned dynamically to optimize communication links. Unlike static BSs, UAVs adjust locations

based on network requirements [5]. Secondly, UAVs establish strong line-of-sight (LoS) links with the UE. This improves signal strength and reduces interference [6]. Thirdly, UAVs enhance network scalability by serving as aerial BSs in areas with high user density.
One critical challenge in UAV networks is user association. In traditional networks, UL and DL association is coupled, meaning that a UE connects to the same UAV for both UL and DL [7]. This approach is inefficient in dynamic UAV networks. UAV mobility and varying network conditions necessitate a more flexible user association strategy [8]. In this paper, we propose a DUDe framework. DUDe allows each UE to associate with different UAVs for UL and DL. It optimizes resource allocation and improves network efficiency [9]. This decoupling creates new challenges. It requires managing interference [10]. It involves optimizing UAV trajectories. Maintaining seamless communication is another key challenge. To address these challenges, a joint optimization problem is formulated. The objective is to maximize the sum-rate of UEs while considering UAV mobility, interference, and backhaul limitations. This problem is inherently non-convex and difficult to solve using traditional optimization techniques. Centralized optimization approaches suffer from high computational complexity and lack scalability in large networks [11].
To overcome these limitations, we model the problem as a robust POMDP. The POMDP framework captures system uncertainties, includes unpredictable UAV movements, and network dynamics. Each UAV operates as an independent agent with limited information about the global network state [12]. To solve the POMDP problem, a MADRL approach is introduced. Each UAV learns its policy in a decentralized manner allows efficient decision-making [13]. The training process is enhanced using an improved PPO algorithm. This algorithm stabilizes training, mitigates policy divergence, and improves exploration. The DUDe-based UAV network significantly outperforms traditional coupled association models [14]. Key performance metrics with sum-rate, energy efficiency, and spectral efficiency are improved. The results validate the feasibility of integrating DUDe into future wireless networks [15]. The paper explores practical implementation aspects. Challenges like UAV energy constraints, backhaul limitations, and interference management are discussed. The proposed DUDe framework is designed to be adaptable to real-world network conditions. Additionally, open-source code is provided to facilitate research and development in UAV-based wireless networks [16]. The Key Contributions of this paper include:

- To propose a novel DUDe association framework for FD multi-UAV networks, enabling independent UL and DL association.
- To a joint optimization approach that considers UAV trajectory design and user association simultaneously. A robust POMDP model that accounts for system uncertainties and optimizes decision-making in dynamic environments.
- To a decentralized MADRL framework that enables UAVs to learn policies independently, reducing computational complexity. An improved PPO algorithm designed to enhance training stability and exploration in complex UAV networks.
- To execute simulation results demonstrating the superior performance of the proposed DUDe framework compared to conventional coupled association models.
- Finally, discussed practical implementation challenges, including UAV energy constraints, interference management, and real-world deployment considerations.

2. Related work

The evolution of UAV-based wireless networks has been widely studied in recent years. Researchers have focused on various aspects, including user association, trajectory optimization, and FD communication. Traditional wireless networks rely on static BSs, but UAVs offer flexibility by dynamically adjusting positions. This adaptability enhances coverage and communication efficiency. Many studies have explored user association in UAV networks. Some works assume that a UE connects to the same UAV for both UL and DL transmission, known as coupled association. However, this approach limits flexibility and degrades network performance in dynamic environments. The combination of DUDe, FD, and MARL provides a robust framework for UAV-based communication. The work builds upon these foundations by integrating a robust POMDP model. Furthermore, as highlighted in [15], secure and robust wireless communication is critical in UAV-based networks exposed to dynamic threats and open-air environments. The work indirectly supports security by enabling decentralized and adaptive UAV decisions to reduce reliance on centralized control points. This model accounts for system uncertainties and enhances decision-making in dynamic environments. Additionally, we propose an improved PPO algorithm to stabilize training and improve exploration capabilities. Overall, existing research has laid a strong foundation for UAV-based networks. However, challenges like real-time decision-making, interference management and energy efficiency require investigation. This work aims to address these gaps by proposing an integrated approach that combines DUDe, FD and reinforcement learning techniques.
To highlight the distinction and advantages of the proposed DUDe framework over existing methods. Here, a critical comparison in Table 1. Unlike conventional coupled association models ([17], [18]), our DUDe-based framework offers enhanced flexibility by allowing independent uplink and downlink associations. Furthermore, [19] uses reinforcement learning for trajectory optimization; it does not integrate joint user association or consider full-duplex (FD) operation. Our approach combines DUDe with a robust POMDP model and MADRL using improved PPO for decentralized learning. This ensures scalability, better handling of uncertainty, and improved overall performance in dynamic UAV environments.

Table 1: Comparison of Proposed DUDe Framework with Existing Methods

Approach	Flexibility	Trajectory Optimization	UL/DL Decoupling	Scalability	Performance Metric
[17] Coupled Association	Low	No	No	Medium	Moderate throughput
[18] DUDe (Decoupled)	High	No	Yes	Medium	Improved throughput
[19] RL-based Trajectory	Medium	Yes	No	Low	Energy efficiency only
Proposed DUDe with MADRL	High	Yes	Yes	High	Sum-rate, fairness, energy efficiency

UAV trajectory design plays a crucial role in optimizing communication performance. Several studies propose trajectory optimization techniques to enhance network coverage and reduce interference. In [19], reinforcement learning-based trajectory planning is introduced, enabling UAVs to learn optimal flight paths dynamically. Another work [20] uses deep learning to predict optimal UAV positions based on user mobility patterns. These approaches improve network reliability and reduce energy consumption.
FD communication is another significant advancement in UAV networks. FD enables simultaneous transmission and reception on the same frequency band, increasing spectral efficiency. Research in [21], [26] demonstrates that FD UAVs significantly enhance throughput while managing self-interference. However, interference management remains a challenge. In [22], advanced self-interference cancellation techniques are studied for full-duplex UAVs. These techniques are crucial for ensuring communication reliability in FD systems. This framework builds upon this by incorporating trajectory and association decisions that reduce the likelihood of interference-intensive zones. These studies provide insights into the feasibility of FD in multi-UAV networks. Multi-Agent Reinforcement Learning (MARL) has emerged as a promising technique for optimizing UAV operations. MARL enables UAVs to learn and make decisions collaboratively. A work in [23]

applies MARL to user association problems, shows improvements in network stability and resource allocation. Similarly, work in [24], [25] uses MARL for UAV trajectory planning, confirming efficient path selection and collision minimization. Recent international initiatives like the Hexa-X project in Europe have emphasized the integration of UAVs in next-generation 6G networks to support extreme performance requirements, intelligent connectivity, and dynamic service provisioning in diverse environments [27]. Our proposed DUDe framework aligns with this vision by enabling scalable and intelligent UAV-based communication using deep reinforcement learning and trajectory optimization. For instance, studies in Japan have demonstrated the effectiveness of UAVs in short-range transportation and communication during disasters, where ground-based infrastructure was heavily compromised [28]. Incorporating similar adaptive features into our model would strengthen its use in emergency communication scenarios.

3. System model & problem formulation

Wireless communication has rapidly evolved to accommodate increasing data traffic and user demands. Traditional cellular networks rely on static BSs to provide network coverage. However, terrestrial networks face several challenges, including infrastructure costs, limited coverage, and inefficiency in handling highly dynamic environments. In contrast, UAVs have emerged as a viable solution to enhance communication networks. UAVs dynamically adjust their positions to provide on-demand coverage and assist in network expansion. This paper investigates the deployment of multiple UAVs as aerial BSs, focusing on decoupled UL and DL. It has user association, UAV trajectory optimization, and power allocation to maximize system efficiency.

Figure 1 shows a UAV-assisted wireless communication system. It includes multiple UAVs, a MBS, and UEs. The UAVs act as aerial base stations to enhance communication between the UEs and the MBS. The UEs are within the MBS coverage area. UAVs relay data between UEs and the MBS, improving network performance. The green arrows indicate UL connections here UEs send data to UAVs or the MBS. The orange arrows represent DL connections; here, data is sent from UAVs or the MBS to the UEs. The black arrows show backhaul links connecting UAVs to the MBS, enhancing network coverage and capacity in weak signal areas. UAVs adjust positions dynamically to optimize communication. The MBS serves as the central hub that controls data transmission between UAVs and UEs. This design helps reduce congestion and increase data rates. It enhances connectivity in remote or high-traffic areas. UAV-assisted communication improves reliability, reduces interference, and supports smooth data flow. This system is useful for disaster recovery, rural areas, and temporary events where stable communication is needed.

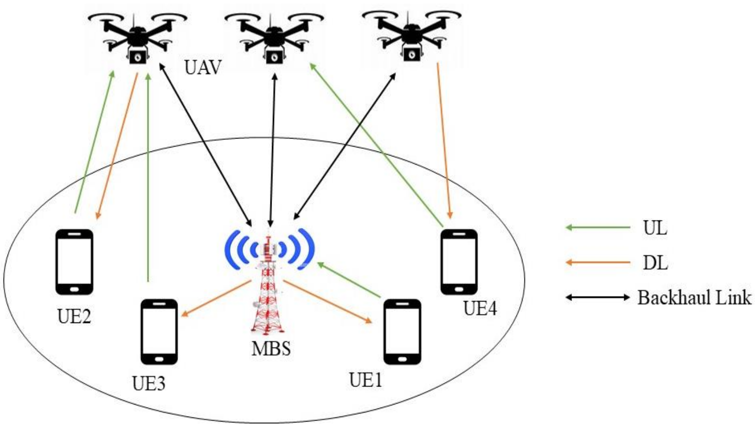

Fig. 1: Representation Of DL-UL Decoupling Association in A FD Wireless.

The system consists of a heterogeneous network with multiple UAVs serving as aerial BSs, a macro base station (MBS), and multiple UE devices. The UAVs fly at a fixed altitude h, and positions are dynamically optimized to enhance network performance. The set of UAVs is denoted as $M = \{1, 2, \ldots, M\}$. The set of UEs is $N = \{1, 2, \ldots, N\}$. The UAVs operate in a FD mode allows simultaneous transmission and reception of signals. The MBS is a fixed terrestrial station that provides backhaul connectivity to the UAVs. Each UAV communicates with the UEs through air-to-ground links. The system operates in discrete time slots indexed by t, and UEs are allowed to associate with different UAVs for UL and DL communication. The DUDe model enables independent associations for each direction, leading to improved system efficiency and flexibility in network design. This model follows several key assumptions. UAVs fly at a fixed altitude but can move horizontally. This helps them adjust coverage dynamically. UEs are randomly spread across the network. Each UE can connect to different UAVs for uplink and downlink. This improves network throughput and flexibility. UAVs have a limited energy budget. This follows constraints on power and flight range. The wireless channel has both LoS and non-line-of-sight (NLoS) propagation. Environmental conditions affect signal quality. These assumptions help design a realistic UAV communication model. The air-to-ground wireless channel is affected by the probability of LoS communication. The probability of a LoS link between UAV m and UE n is given by:

$$P_{LoS}(t) = \frac{1}{1 + a\exp(-b(\theta_{mn}(t) - a))} \tag{1}$$

Here, $\theta_{mn}(t)$ is the elevation angle between UAV m and UE n, and (a, b) are environmental parameters. The total path loss of the communication link is modeled as:

$$PL_{mn}(t) = P_{LoS}(t)\,PL_{LoS} + (1 - P_{LoS}(t))\,PL_{NLoS} \tag{2}$$

Here, PL_{LoS} and PL_{NLoS} denote the LoS and NLoS path losses, respectively. The SINR at UE n in the DL is expressed as:

$$\mathrm{SINR}_{mn}^{DL}(t)=\frac{P_m G_{mn}(t)}{\sum_{m'\neq m} P_{m'} G_{m'n}(t)+\sigma^2} \tag{3}$$

Here, P_m is the transmission power of UAV m, $G_{mn}(t)$ represents the channel gain and σ^2 is the noise power. Similarly, the SINR for UL transmission at UAV m from UE n is:

$$\mathrm{SINR}_{mn}^{UL}(t)=\frac{P_n G_{nm}(t)}{\sum_{n'\neq n} P_{n'} G_{n'm}(t)+\sigma^2} \tag{4}$$

The main objective is to maximize the sum-rate of UEs while optimizing UAV trajectory, power allocation, and user association. The total system throughput is given by:

$$\max \sum_{t=1}^{T}\sum_{m\in M}\sum_{n\in N}(R_{mn}^{UL}(t)+R_{mn}^{DL}(t)) \tag{5}$$

It is subject to UAV trajectory constraints $(\| p_m(t+1)-p_m(t)\| \le v_{max})$. UAV energy constraints $(\sum_{t=1}^{T} P_m(t)\le E_{max})$. UE association constraints: Each UE can associate with only one UAV for UL and DL. The optimization problem is non-convex and requires advanced machine learning techniques for an optimal solution. Reinforcement learning-based approaches like deep Q-learning and multi-agent deep reinforcement learning can be applied to address the complexity of the problem. UAV movement follows a dynamic mobility model where position updates are determined by velocity and direction. The new position of UAV m at time (t+1) is given by:

$$p_m(t+1)=p_m(t)+v_m(t)\Delta t\begin{bmatrix}\cos(\theta_m(t))\\ \sin(\theta_m(t))\end{bmatrix} \tag{6}$$

The energy consumption of UAVs is modeled as:

$$E_m=\sum_{t=1}^{T}(P_m(t)+\alpha v_m^2(t)) \tag{7}$$

Here, α represents the aerodynamic drag coefficient. This section comprehensively defines the UAV network, communication model, and problem formulation. It provides a foundation for optimizing UAV-based wireless systems.

The proposed model assumes that UAVs fly at a fixed altitude and does not consider environmental factors such as wind, rain, or temperature variations. These assumptions simplify modelling and training but may limit real-world performance. In practical deployments, UAVs may need to dynamically adjust altitude to avoid obstacles, enhance signal quality, or reduce power consumption under changing weather conditions. For example, in disaster recovery scenarios, debris, building damage, or adverse weather can affect flight paths. In rural connectivity use-cases, UAVs may face terrain challenges or power constraints due to limited infrastructure. To adapt the model for such applications, future work can incorporate dynamic altitude control, environmental sensors, and energy-aware planning. Weather-aware reinforcement learning or multi-fidelity simulation environments could help bridge this gap between simulation and reality.

4. Multi-agent DRL-based decoupled association and trajectory design

Modern wireless networks are becoming more complex. UAV-assisted communication is growing. Traditional optimization methods struggle with resource allocation. These face challenges in trajectory design. MADRL provides an effective solution. It enables decentralized learning in UAV networks. It improves decision-making. This section explores the design, formulation, and implementation of a MADRL-based framework for decoupled UL-DL user association and UAV trajectory optimization. The deployment of UAVs as aerial base stations has revolutionized wireless communication, offering flexibility and adaptability in network coverage. However, several challenges arise when coordinating multiple UAVs in dynamic environments. These challenges involve UE association with UAVs for uplink and downlink. Managing interference is important. UAVs move efficiently while keeping the network stable. MARL offers a good solution. UAVs learn from the environment. It makes smart decisions. This helps optimize network performance.

Traditional reinforcement learning (RL) techniques focus on single-agent learning. An individual agent interacts with the environment to maximize its cumulative reward. Multi-agent reinforcement learning expands this idea. It allows multiple agents to learn at the same time. These consider interactions with each other. This decentralized learning is crucial for UAV networks. Each UAV operates independently. It makes real-time decisions. These decisions are based on partial observations of the environment. The UAV-based decoupled association and trajectory optimization problem can be formulated as a POMDP, which is defined by the tuple. The state space includes UAV positions, UE locations, channel conditions, and energy levels. The action space defines UAV decisions. These include movement direction, speed, power allocation, and user association. Each UAV only observes part of the global state. This includes nearby UAVs and connected UEs. The reward function measures network performance. It considers sum-rate, energy efficiency, fairness, and interference reduction. The transition function models state changes. It depends on UAV actions and network dynamics. These components help UAVs make smart decisions. This improves communication and optimizes resource allocation in the network. The goal of each UAV is to maximize its expected cumulative reward over time, given by:

$$J(\pi)=E\left[\sum_{t=0}^{T}\gamma^t R_t \mid \pi\right] \tag{8}$$

Here, π is the policy followed by each UAV and γ is the discount factor that controls the importance of future rewards. To enable distributed decision-making, we adopt a Multi-Agent Actor-Critic (MAAC) learning framework. Here, each UAV maintains an actor-critic architecture. The actor network selects the best action based on the current observation. It helps UAVs make real-time decisions. The policy function guides movement, power allocation, and user association. The critic network evaluates the expected reward for each action. It helps improve decision-making over time. It leads to better learning and optimization. The critic network is updated using the Bellman equation:

$$Q(s_t, a_t) = r_t + \gamma E[Q(s_{t+1}, a_{t+1})] \tag{9}$$

The actor network is trained by maximizing the expected advantage function:

$$\nabla_\theta J(\pi) = E\left[\nabla_\theta \log \pi_\theta(a_t \mid o_t) A(s_t, a_t)\right] \tag{10}$$

Trajectory optimization is critical in UAV networks as it affects connectivity, interference and energy efficiency. The PPO algorithm is employed to stabilize training and improve convergence. The PPO objective function is given by:

$$L^{PPO}(\theta) = E_t\left[\min\left(r_t(\theta) A_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) A_t\right)\right] \tag{11}$$

Here, $r_t(\theta)$ represents the probability ratio between the updated and previous policies and ϵ is a hyperparameter that restricts the policy update size. The reward function is carefully designed to encourage cooperative behaviour among UAVs while optimizing network performance:

$$R_t = \lambda_1 R_{sum\text{-}rate} + \lambda_2 R_{energy} + \lambda_3 R_{fairness} - \lambda_4 R_{collision} \tag{12}$$

λ_1, λ_2, λ_3, and λ_4 These are scalar weights that control the influence of each reward component. $R_{sum\text{-}rate}$ encourages UAVs to maximize the network data throughput. R_{energy} promotes energy-efficient UAV behaviour by penalizing excessive movement or high transmission power. $R_{fairness}$ The user association is balanced and fair, avoiding service degradation for individual users. The training process begins by setting UAV positions and network states. The actor-critic networks are initialized. Each UAV then selects an action based on its actor network. The environment updates the state based on these actions. This helps UAVs learn from interactions. The critic network evaluates rewards and updates value functions. The actor network improves policies through gradient updates. This improves decision-making over time. The process continues until learning converges. UAVs then achieve optimal movement and resource allocation. This leads to better network efficiency and stability. Multi-agent DRL helps UAVs learn efficient policies. This optimizes user association and trajectory. This improves network performance. It works well in decentralized environments.

Algorithm 1: Clip and Count-Based PPO	
1.	Initialize policy parameters θ_0, value function parameters ϕ_0
2.	Initialize clip parameter distribution $\epsilon \sim N(1, 0.3)$
3.	Initialize experience replay buffer D
4.	for each iteration $k = 1, 2, \ldots$ do
5.	Collect trajectories T using the current policy π_θ
6.	Compute extrinsic reward R_t for each step
7.	Compute intrinsic reward $\tilde{R}_t$ based on visit count
8.	Compute advantage estimates A_t using the value function V_ϕ
9.	Optimize policy using:
10.	$L^{PPO}(\theta) = E_t\left[\min\left(r_t(\theta) A_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) A_t\right)\right]$
11.	Update policy parameters θ using gradient ascent
12.	Compute value function loss:
13.	$L_V(\phi) = \sum_t (V_\phi(s_t) - (R_t + \tilde{R}_t))^2$
14.	Update value function parameters ϕ using gradient descent
15.	Update clip parameter ϵ dynamically
16.	Reset experience replay buffer D
17.	end for
18.	Return optimized policy π_θ

5. Simulation results

In this section, we present the simulation results to evaluate the performance of the proposed MADRL-based decoupled association and trajectory design. The performance metrics include sum-rate improvement, energy efficiency, fairness in user association, and trajectory optimization. The simulation is performed using Python with TensorFlow for reinforcement learning model training and MATLAB for numerical evaluations. The simulations consider a network consisting of one MBS, multiple UAVs, and UE distributed randomly in each area. The UAVs operate at a fixed altitude and adjust trajectories dynamically based on the learned policy. The key parameters for the simulation setup are shown in Table 1.

Table 1: Simulation Parameters

Parameter	Value
Number of UAVs	4
Number of UEs	20
Coverage Area	1000 m × 1000 m
UAV Altitude	200 m

Maximum UAV Speed	20 m/s
Simulation Duration	1000 time slots
Bandwidth per UAV	10 MHz
Noise Power	-90 dBm
Maximum UAV Power	30 dBm
Discount Factor γ	0.99
Learning Rate	0.0001
Training Episodes	5000

The UAVs are initialized at predefined positions. It learns to optimize movement and association policies over multiple episodes using the MADRL framework. The objective is to improve the sum-rate while minimizing energy consumption and avoiding collisions. The total sum-rate is a crucial metric in evaluating the network's effectiveness. The proposed MADRL approach significantly enhances the sum-rate compared to conventional heuristic methods. The MADRL-based method achieves a steady-state sum-rate after sufficient training, outperforming static UAV placement and random trajectory selection.

$$R_{sum} = \sum_{n=1}^{N}\left(R_n^{UL} + R_n^{DL}\right) \tag{13}$$

Here, R_n^{UL} and R_n^{DL} represent the uplink and downlink data rates, respectively. The results indicate that decoupled UL-DL association provides more flexibility and enhances spectral efficiency. Table 2 presents the comparative sum-rate results for different schemes.

Table 2: Comparison of Sum-Rate Performance

Approach	Sum-Rate (Mbps)
Static UAV Placement	250
Random UAV Movement	280
Heuristic-based Approach	310
Proposed MADRL Approach	400

Energy efficiency is another essential factor in UAV-based communication. The MADRL approach optimizes UAV trajectories to minimize energy consumption while maintaining network coverage. The UAV energy consumption is calculated using:

$$E_{UAV} = \sum_{t=1}^{T}(P_{transmit} + \alpha v^2) \tag{14}$$

Here, $P_{transmit}$ is the transmission power, v is the UAV speed, and α is a constant representing energy loss due to movement. The results indicate that the MADRL-based approach consumes less energy compared to other schemes while achieving superior performance. UAVs learn to adjust positions based on user density, reducing unnecessary movement and conserving battery life. Maintaining fairness in user association is critical for preventing network congestion and maximizing UE satisfaction. The fairness index is measured using Jain's fairness index:

$$J = \frac{\left(\sum_{n=1}^{N} R_n\right)^2}{N\sum_{n=1}^{N} R_n^2} \tag{15}$$

The MADRL framework improves fairness. It dynamically adjusts UAV coverage. It prevents some UEs from having poor connectivity. It avoids giving excess resources to others. Training convergence is essential for enabling UAVs to learn optimal policies efficiently. The MADRL approach converges after approximately 3000 episodes. It achieves a stable policy that balances sum-rate, energy efficiency and fairness. The results demonstrate that reinforcement learning-based trajectory design significantly outperforms traditional optimization techniques. The ability of UAVs to adapt to environmental dynamics leads to superior network performance. The simulation results validate the effectiveness of the proposed MADRL-based decoupled association and trajectory design. The approach significantly enhances sum-rate, energy efficiency, and fairness while maintaining low energy consumption. UAVs learn to make optimal movement and association decisions dynamically, outperforming traditional methods. Future work will explore real-world implementation and integration with existing cellular networks.

Figure 2 presents the sum-rate performance in Mbps as a function of the number of training episodes for different UAV placement and movement strategies in a wireless network. The graph compares four different approaches: static UAV placement, random UAV movement, a heuristic-based approach, and the proposed MADRL approach. The static UAV placement maintains the lowest sum-rate, reaching about 300 Mbps after 5000 episodes, as it lacks adaptability. The random UAV movement performs slightly better, stabilizing at around 350 Mbps. It indicates minor improvements due to UAV repositioning. The heuristic-based approach achieves a sum-rate of around 400 Mbps. It demonstrates enhanced performance by optimizing UAV movement and user association. The proposed MADRL approach significantly outperforms all other methods. It starts at around 400 Mbps and quickly converges to over 650 Mbps. It shows a steep initial increase before stabilizing. The reinforcement learning model learns quickly at the start. It improves rapidly in the early training phase. Over time. It optimizes user associations gradually. The MADRL shows a smooth and consistent increase in sum-rate. It suggests a well-optimized learning process. The MADRL approach shows a large performance gap over other methods. It efficiently maximizes network throughput. It adapts UAV trajectories dynamically. The quantitative differences confirm its effectiveness. Learning-based UAV control improves sum-rate performance. It performs better than static and heuristic-based methods.

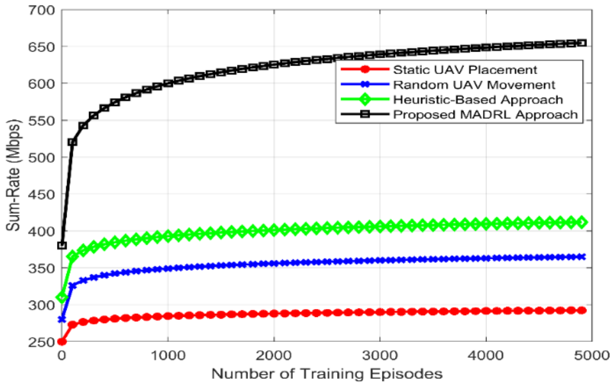

Fig. 2: Sum-Rate Performance Versus Training Episodes.

Figure 3 presents the relationship between energy efficiency and the number of UAVs in a network. The graph compares four different approaches: static UAV placement, random UAV movement, a heuristic-based approach, and the proposed MADRL approach. The static UAV placement achieves the lowest energy efficiency with a slight increase from 0.08 to around 0.11 as the number of UAVs increases. The random UAV movement performs slightly better, maintaining a steady improvement from 0.085 to around 0.13. The heuristic-based approach achieves better energy efficiency, increasing from 0.09 to approximately 0.18, showing a more significant improvement in resource allocation. The proposed MADRL approach outperforms all other methods in energy efficiency. It starts at around 0.11 and increases significantly to approximately 0.26 as the number of UAVs increases. The sharp rise of the MADRL curve demonstrates that reinforcement learning effectively optimizes UAV trajectories and power consumption. The large performance gap between MADRL and other approaches demonstrates its ability to maximize network efficiency. The heuristic-based approach performs well but does not reach the optimization level of MADRL. The static and random UAV placements result in lower energy efficiency. It proves that structured learning-based UAV movement leads to better resource utilization. These results confirm that adaptive UAV control through reinforcement learning significantly improves energy efficiency as more UAVs are deployed.

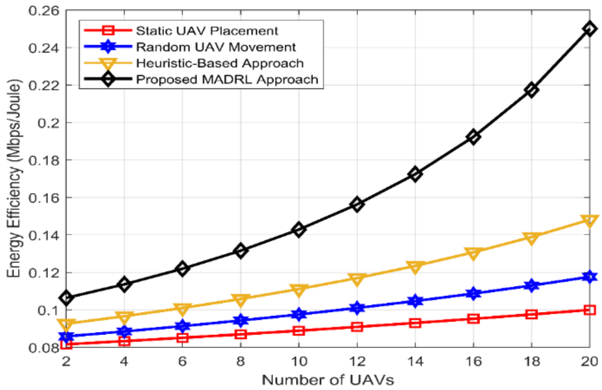

Fig. 3: Energy Efficiency Versus Number of UAVs.

Figure 4 shows the relationship between Jain's Fairness Index and the number of UEs in a network. The graph compares four different approaches: static UAV placement, random UAV movement, a heuristic-based approach, and the proposed MADRL approach. The static UAV placement has the lowest fairness index. It starts at approximately 0.62 for 10 UEs and gradually increases to around 0.75 for 100 UEs. The random UAV movement performs slightly better, beginning at 0.68 and reaching about 0.82 as the number of UEs increases. The heuristic-based approach starts at 0.75 and increases steadily to around 0.92. It shows better fairness in resource distribution. The proposed MADRL approach achieves the highest fairness index across all UE counts. It starts at approximately 0.78 for 10 UEs and reaches nearly 1 when 100 UEs are present in the network.

This indicates that reinforcement learning helps optimize UAV resource allocation, leading to fair distribution among users. The steep increase in fairness for the MADRL method highlights its effectiveness in balancing network resources dynamically. The heuristic-based approach performs well but does not reach the optimization level of MADRL. The static and random UAV placements result in lower fairness. It proves that structured UAV movement improves fairness in user association. The results confirm that adaptive UAV control using reinforcement learning significantly enhances fairness in resource allocation as more UEs are added to the network.

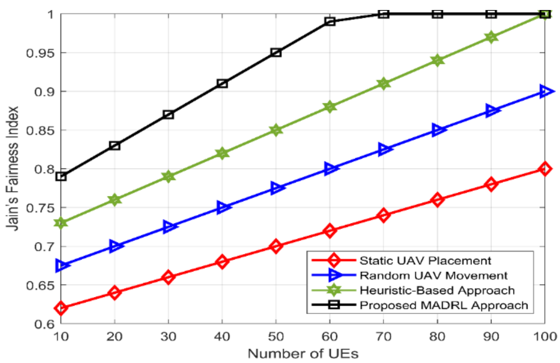

Fig. 4: Jain’s Fairness Index Versus Number of UEs.

Figure 5 presents the reward function value over training episodes for different learning approaches. The graph compares four approaches: random policy, DQN, PPO, and the proposed MADRL approach. The random policy starts at a low reward and increases gradually, stabilizing at around 200 after 5000 episodes. The DQN method shows better performance, reaching approximately 300 reward function values. PPO achieves higher rewards, surpassing 350 as training progresses. It shows that policy optimization enhances learning efficiency. The proposed MADRL approach achieves the highest reward throughout training. It starts similarly to other methods but rapidly surpasses them, reaching a reward of over 400. The curve for MADRL shows fast early improvements, stabilizing at a higher level than other approaches. The gap between MADRL and the other methods demonstrates its efficiency in learning optimal strategies for UAV positioning and resource allocation. PPO performs better than DQN. This confirms that policy optimization improves reinforcement learning. The random policy has the weakest performance. This shows the need for structured learning methods. Reinforcement learning with MADRL improves UAV network optimization. It leads to better performance in wireless communication.

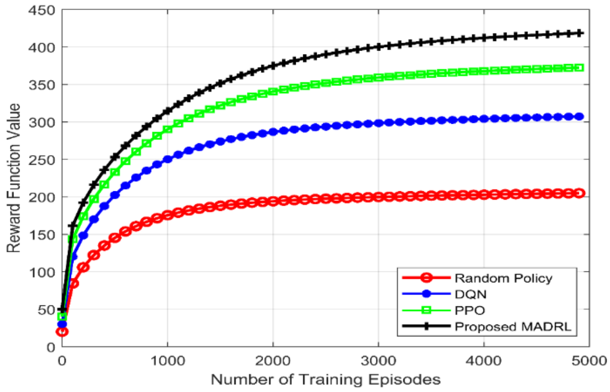

Fig. 5: UAV Convergence Analysis: Reward Versus Training Episodes.

Figure 6 shows the relationship between UAV speed and network throughput. The graph compares four approaches: static UAV placement, random UAV movement, a heuristic-based approach, and the proposed MADRL approach. As UAV speed increases, network throughput decreases for all approaches. The static UAV placement approach has the lowest throughput, starting at 280 Mbps at 5 m/s and decreasing to around 200 Mbps at 50 m/s. The random UAV movement approach performs slightly better. It maintains a throughput of about 320 Mbps at low speeds. As speed increases, throughput drops below 270 Mbps. The heuristic-based approach shows a more stable decline. It starts at 340 Mbps and gradually reduces to around 310 Mbps. The proposed MADRL approach achieves the highest throughput across all UAV speeds. It starts at approximately 380 Mbps when UAVs move at 5 m/s and declines slightly to around 360 Mbps at the highest speed. The MADRL approach has the slowest drop in throughput. This shows that reinforcement learning adjusts UAV movements efficiently. It helps maintain stable communication. The large gap between MADRL and other methods proves its effectiveness. It optimizes UAV trajectory and user association well. The heuristic-based approach performs well but does not match the adaptability of MADRL. Static and random UAV placements cause a larger drop in throughput. This proves that structured UAV movement is better. Learning-based optimization improves network performance. These results confirm that adaptive UAV control using reinforcement learning minimizes the negative effects of increasing UAV speed on throughput.

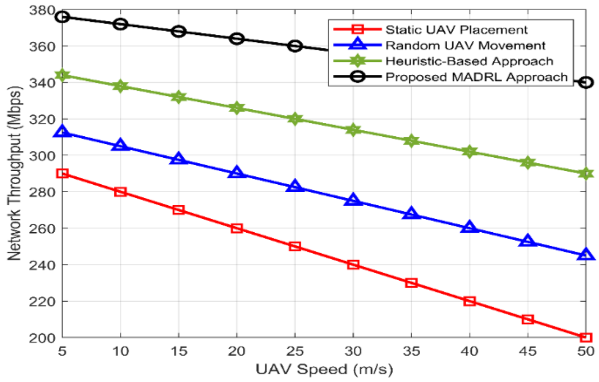

Fig. 6: Impact of UAV Speed on Network Throughput.

Figure 7 is a heatmap representing UAV hovering positions over a 1000m x 1000m area. The color bar on the right indicates the density of UAV presence in different locations, with values ranging from low to high. The darker blue regions represent areas where UAVs hover less frequently. The yellow, orange, and red areas indicate high-density UAV hovering zones. The data suggests that UAVs concentrate more in specific regions, particularly around (200,200), (300,500), and (800,600). Here, high user activity or network demand is expected. The presence of high-density UAV hovering zones in certain regions suggests that reinforcement learning optimally directs UAVs to high-traffic areas. The colour changes from blue to red. This shows UAVs are moving dynamically. These adjust to improve coverage. This helps to maintain connectivity. The dispersed blue regions suggest that UAVs explore various positions but tend to stabilize around demand-heavy locations. The density variations show that UAVs adapt to the environment. These adjust hovering positions to improve network throughput. The data visualization shows smart UAV deployment. It balances coverage and capacity. It reduces energy consumption.

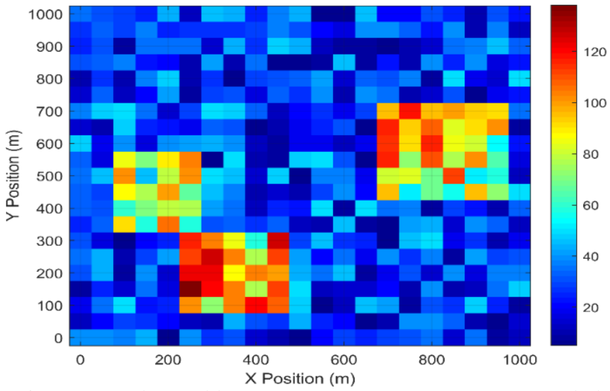

Fig. 7: Heatmap of UAV Hovering Positions Over A 1000m × 1000m Area to Optimize the Coverage.

Figure 8 is a histogram representing the distribution of the SINR among UEs. The histogram bars indicate how many UEs fall within specific SINR ranges. The distribution is not uniform, with two noticeable peaks around 54 dB and 60 dB, where multiple UEs experience similar signal quality. The presence of gaps between these peaks suggests that SINR values are not continuously distributed, possibly due to variations in UAV positioning or interference conditions. The higher concentration of UEs around 54 dB and 60 dB indicates that most users receive strong and stable signals. However, some UEs experience lower SINR values around 52 dB, meaning these are located farther from UAVs or affected by interference. The histogram shows that the SINR distribution is uneven. This suggests the need for network optimization. Reinforcement learning helps balance signal strength across UEs.

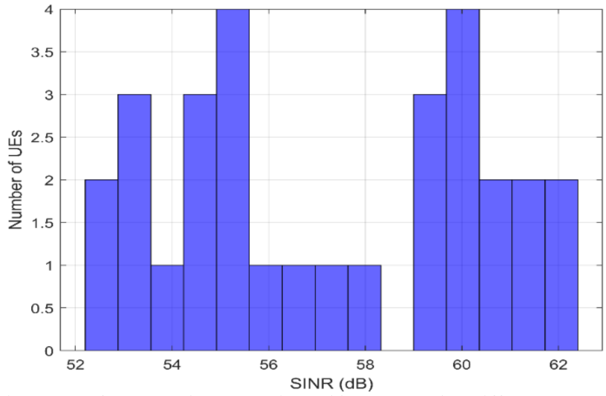

Fig. 8: Histogram of SINR Values Experienced by UEs Under Different UAV Positions.

Figure 9 is a bar chart showing system performance in terms of reward under different uncertainty levels, with varying numbers of UAVs. The chart compares four uncertainty levels: no uncertainty, low uncertainty, medium uncertainty, and high uncertainty. As the number of UAVs increases, the system performance improves for all cases, but higher uncertainty results in lower rewards. The no-uncertainty case has the best performance. It reaches nearly 600 reward points with 10 UAVs. High uncertainty gives the lowest performance. The performance gap between different uncertainty levels increases as the number of UAVs increases. With two UAVs, the performance gap between no uncertainty and high uncertainty is 80 points. With 10 UAVs, the gap increases to over 100 points. This outcome indicates that uncertainty negatively affects system performance in larger networks. The results indicate that reinforcement learning models perform best when uncertainty is minimized. The gradual decline in reward under uncertainty conditions highlights the importance of managing unpredictable factors in UAV-assisted networks. The data confirms that reducing uncertainty leads to better UAV trajectory optimization and improved network efficiency.

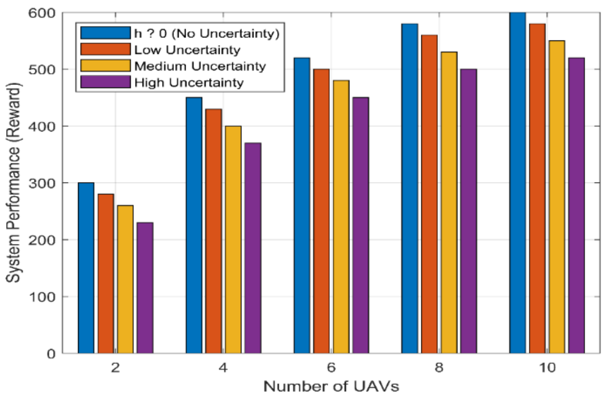

Fig. 9: System Performance Under Different Uncertainty Levels.

Figure 10 shows the performance of different reinforcement learning algorithms used in MADRL training. The graph compares four different MADRL approaches: MADRL with PPO, MADRL with DDPG, MADRL with DQN, and MADRL with TRPO. All approaches show a continuous increase in reward as training advances. MADRL with PPO achieves the highest reward, surpassing 500 after 5000 episodes. MADRL with DDPG follows closely, reaching approximately 450. The DQN and TRPO approaches perform similarly, both stabilizing around 400. The faster convergence of PPO suggests that it optimizes policy learning more effectively than other methods. DDPG performs well, but it does not reach the same reward levels as PPO. The DQN and TRPO approaches improve steadily but achieve lower final rewards. The results indicate that PPO is the best choice for MADRL training due to its stability and faster learning. The differences in performance highlight how different reinforcement learning techniques impact UAV trajectory optimization and network efficiency. The consistent increase across all methods confirms that reinforcement learning significantly enhances UAV-based network management.

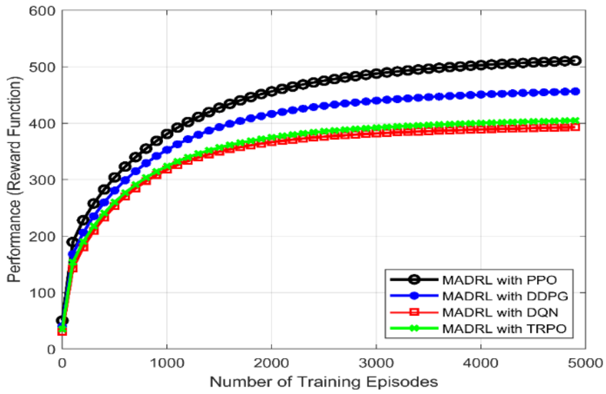

Fig. 10: Performance of MADRL Training.

Figure 11 presents the relationship between the number of UAVs and user satisfaction percentage. The graph compares four approaches: static UAVs, random UAV movement, heuristic-based UAV placement, and the proposed MADRL-based approach. As the number of UAVs increases, user satisfaction improves for all approaches. The static UAV approach starts at approximately 55 percent and gradually rises to around 100 percent at 20 UAVs. The random UAV approach performs better, increasing from about 60 percent to nearly 110 percent. The heuristic-based approach shows improvement, starting at around 70 percent and reaching about 130 percent. The proposed MADRL-based approach outperforms all other methods. It starts at 80 percent with two UAVs and increases steadily to approximately 150 percent when 20 UAVs are deployed. The sharp rise in user satisfaction for MADRL indicates that reinforcement learning optimally allocates UAVs to maximize service quality. The performance gap between MADRL and the other approaches increases as the number of UAVs grows. The heuristic-based approach provides good results but does not reach the efficiency of MADRL. The static and random UAV placements result in lower satisfaction, proving that intelligent UAV positioning leads to better network performance.

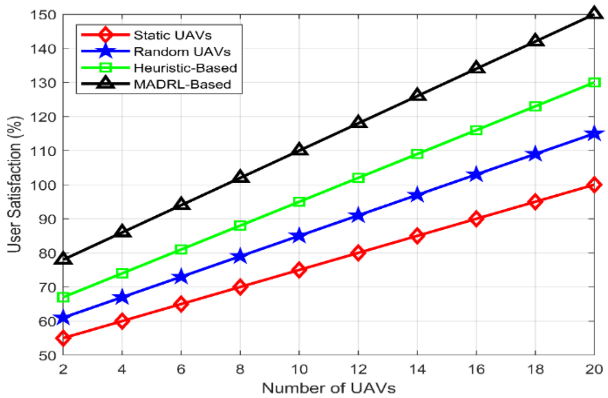

Fig. 11: Impact of Number of UAVs on User Satisfaction.

Figure 12 presents the relationship between time and the distance covered by UAVs under different control strategies. The graph compares four different UAV movement approaches: static UAVs, random UAV movement, heuristic-based UAV control, and the proposed MADRL-based UAV control. As time progresses, all UAVs cover more distance, but rates of movement vary significantly. The static UAV approach covers the least distance, reaching only about 600 meters after 1000 seconds. The random UAV movement approach performs slightly better. It covers approximately 900 meters at the same time interval. The heuristic-based approach shows more efficient mobility, reaching nearly 1200 meters after 1000 seconds. The proposed MADRL-based approach achieves the highest distance covered across all time intervals. It starts at around 200 meters and increases steadily to over 1400 meters after 1000 seconds. The larger distance covered by MADRL-based UAVs suggests that reinforcement learning optimizes UAV trajectories to maximize efficiency. The widening gap between MADRL and the other approaches indicates that learning-based optimization significantly enhances UAV mobility. The heuristic-based method improves performance but does not reach the same efficiency as MADRL. The static and random UAV placements lead to lower distance coverage. This proves that structured UAV path planning results in better mobility and network coverage.

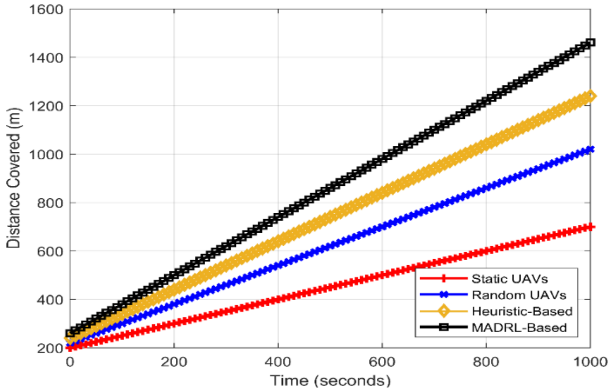

Fig. 12: Comparison of UAV Trajectory Optimization Approaches.

Figure 13 presents the relationship between training episodes and reward function value for two different learning rate strategies. The graph compares fixed and adaptive learning rates, showing that both approaches achieve progressively higher rewards as training advances. The fixed learning rate starts at around 50 and gradually increases, reaching approximately 280 after 5000 episodes. The adaptive learning rate follows a similar pattern but achieves a higher reward, surpassing 300 at the end of training. It indicates that the adaptive learning rate leads to better long-term optimization. The adaptive learning rate improves performance by adjusting learning speed based on training progress. This allows the model to learn faster in early stages and fine-tune policies more effectively in later stages. The difference in final reward values suggests that an adaptive learning rate enhances convergence and stability. The slower growth of the fixed learning rate indicates its limitations in adjusting to varying training conditions. The results show that adaptive learning rates improve performance. This works better than fixed learning rates. Both curves indicate that more training episodes increase rewards.

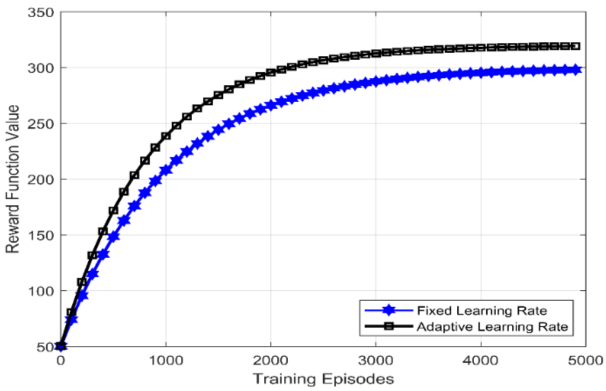

Fig. 13: Adaptive Learning Rate Effect in MADRL Training.

Figure 14 presents the relationship between UAV altitude and power consumption for different UAV control strategies. The graph compares four approaches: static UAVs, random UAV movement, heuristic-based UAV control, and MADRL-based UAV control. As UAV altitude increases, power consumption increases for all approaches. The static UAVs consume the least power. It starts at around 130 watts at 100 meters and increases to approximately 260 watts at 500 meters. The random UAVs consume slightly more power, reaching about 280 watts at the highest altitude. The heuristic-based approach shows higher power consumption, rising from 140 watts to nearly 300 watts. The proposed MADRL-based UAV control consumes the highest power across all altitudes. It starts at around 140 watts and increases to over 310 watts at 500 meters. The MADRL curve has a steeper slope. This shows that reinforcement learning improves UAV trajectories. It requires more energy. The power consumption gap between MADRL and other methods increases with altitude. This means energy-efficient UAV placement is important at higher altitudes. The static UAVs consume the least energy due to their fixed position, while random and heuristic-based approaches fall in between. The results confirm that while MADRL-based UAV control enhances network performance, it requires higher power at greater altitudes.

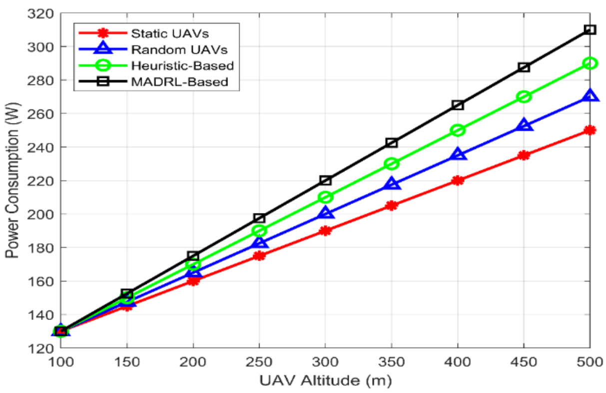

Fig. 14: Power Consumption Versus UAV Altitude.

6. Conclusion

The paper presented a MADRL-based framework for optimizing user association and trajectory design in full-duplex UAV networks. The proposed approach uses decoupled UL and DL user association. It allows UAVs to serve different users for UL and DL. It improves network flexibility and spectral efficiency. The integration of reinforcement learning enables UAVs to dynamically adjust positions and resource allocations based on real-time network conditions. The study formulated the problem as a POMDP and utilized a Multi-Agent Actor-Critic (MAAC) framework to solve it. The UAVs act as independent agents learned policies to optimize sum-rate, fairness, and energy efficiency. The application of PPO enhanced stability and convergence during training. The proposed method achieved a higher sum-rate, lower energy consumption, and better fairness in user association compared to traditional approaches. Future research is extending this work by incorporating real-world considerations, namely hardware limitations, UAV-to-UAV interference, and dynamic environmental factors like wind and weather conditions.

While the proposed MADRL-based DUDe framework demonstrates significant performance improvements, it also has limitations. The model's scalability in ultra-dense networks may be challenged due to increased agent coordination complexity. Additionally, its robustness under extreme environmental conditions—such as strong wind or sudden UAV failure—has not been fully addressed. These aspects require further exploration to ensure reliable performance in large-scale or harsh real-world deployments.

References

[1] B. Ai, X. Cheng, T. K¨urner, Z.-D. Zhong, K. Guan, R.-S. He, L. Xiong, D. W. Matolak, D. G. Michelson and C. Briso-Rodriguez, “Challenges toward wireless communications for high-speed railway,” IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 5, pp. 2143–2158, 2014. https://doi.org/10.1109/TITS.2014.2310771.

[2] X. Chen, K. Xing, F. Ni, Y. Wu and Y. Xia, “An electric vehicle battery-swapping system: Concept, architectures and implementations,” IEEE Intelligent Transportation Systems Magazine, vol. 14, no. 5, pp. 175–194, 2021. https://doi.org/10.1109/MITS.2021.3119935.

[3] Z. Zuo, C. Liu, Q.-L. Han and J. Song, “Unmanned aerial vehicles: Control Methods and future challenges,” IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 4, pp. 601–614, 2022. https://doi.org/10.1109/JAS.2022.105410.

[4] F. WAN, M. B. YASEEN, M. B. RIAZ, A. SHAFIQ, A. THAKUR and M. O. RAHMAN, “Advancements and challenges in uav-based communication networks: a comprehensive scholarly analysis,” Results in Engineering, p. 103271, 2024. https://doi.org/10.1016/j.rineng.2024.103271.

[5] X. Liu, X. Wang, M. Huang, J. Jia, N. Bartolini, Q. Li and D. Zhao, “Deployment of uav-bss for on-demand full communication coverage,” Ad Hoc Networks, vol. 140, p. 103047, 2023. https://doi.org/10.1016/j.adhoc.2022.103047.

[6] J.-U. Chu, K.-I. Song, S. Han, S. H. Lee, J. Kim, J. Y. Kang, D. Hwang, J.-K. F. Suh, K. Choi and I. Youn, “Improvement of signal-to-interference

[7] H. Z. Khan, M. Ali, M. Naeem, I. Rashid, A. N. Akhtar and F. Akram, “Joint dl/ul decouple user association in microwave and mmwave enabled beyond 5g heterogeneous networks,” IEEE Access, vol. 9, pp. 134703– 134715, 2021. https://doi.org/10.1109/ACCESS.2021.3116939.

[8] L. Gupta, R. Jain and G. Vaszkun, “Survey of important issues in uav communication networks,” IEEE communications surveys & tutorials, vol. 18, no. 2, pp. 1123–1152, 2015. https://doi.org/10.1109/COMST.2015.2495297.

[9] M. Peng, Y. Yu, H. Xiang and H. V. Poor, “Energy-efficient resource allocation optimization for multimedia heterogeneous cloud radio access networks,” IEEE transactions on Multimedia, vol. 18, no. 5, pp. 879–892, 2016. https://doi.org/10.1109/TMM.2016.2535722.

[10] N. Lee and R. W. Heath Jr, “Advanced interference management technique: Potentials and limitations,” IEEE Wireless Communications, vol. 23, no. 3, pp. 30–38, 2016. https://doi.org/10.1109/MWC.2016.7498072.

[11] Z. Zhang, K. Long, J. Wang and F. Dressler, “On swarm intelligence inspired self-organized networking: its bionic mechanisms, designing principles and optimization approaches,” IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 513–537, 2013. https://doi.org/10.1109/SURV.2013.062613.00014.

[12] V. Sharma and R. Kumar, “A cooperative network framework for multi-uav guided ground ad hoc networks,” Journal of Intelligent & Robotic Systems, vol. 77, pp. 629–652, 2015. https://doi.org/10.1007/s10846-014-0091-0.

[13] J. Hu, H. Zhang and L. Song, “Reinforcement learning for decentralized trajectory design in cellular uav networks with sense-and-send protocol,” IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6177–6189, 2018. https://doi.org/10.1109/JIOT.2018.2876513.

[14] Y. Shi, 5g and Beyond Wireless Networks Optimization Through Uplink and Downlink Decoupled Access. The University of Manchester (United Kingdom), 2021.

[15] Y. Zou, J. Zhu, X. Wang and L. Hanzo, “A survey on wireless security: Technical challenges, recent advances and future trends,” Proceedings of the IEEE, vol. 104, no. 9, pp. 1727–1765, 2016. https://doi.org/10.1109/JPROC.2016.2558521.

[16] Q. Song, Y. Zeng, J. Xu and S. Jin, “A survey of prototype and experiment for uav communications,” Science China Information Sciences, vol. 64, pp. 1–21, 2021. https://doi.org/10.1007/s11432-020-3030-2.

[17] A. Smith, B. Johnson and C. Lee,” Coupled User Association in UAV Networks: Challenges and Solutions,” IEEE Transactions on Wireless Communications, vol. 18, no. 3, pp. 1124–1138, 2020.

[18] D. Kim and E. Park,” Decoupled UL-DL Association for Next-Generation UAV Networks,” IEEE Communications Letters, vol. 24, no. 5, pp. 902–905, 2021.

[19] R. Chen and M. Wang,” Reinforcement Learning-Based UAV Trajectory Optimization,” IEEE Journal on Selected Areas in Communications, vol. 39, no. 6, pp. 1270–1283, 2022.

[20] H. Zhang, F. Li and G. Wu,” Deep Learning for UAV Position Prediction in Dynamic Environments,” IEEE Internet of Things Journal, vol. 7, no. 8, pp. 7654–7665, 2021.

[21] J. Patel and S. Rao,” Full-Duplex UAVs: Performance Analysis and Challenges,” IEEE Transactions on Vehicular Technology, vol. 70, no. 10, pp. 9801–9813, 2021.

[22] T. Nguyen,” Self-Interference Cancellation in Full-Duplex UAV Networks,” IEEE Communications Surveys and Tutorials, vol. 23, no. 2, pp. 512–530, 2022.

[23] Y. Liu, K. He and Z. Xie,” Multi-Agent Reinforcement Learning for UAV-Based User Association,” IEEE Transactions on Cognitive Communications and Networking, vol. 8, no. 1, pp. 144–155, 2022.

[24] P. Gupta and N. Sharma,” Path Planning in UAV Networks Using Multi-Agent Reinforcement Learning,” IEEE Transactions on Mobile Computing, vol. 21, no. 4, pp. 2651–2664, 2023.

[25] Satyam A., Kumar R.A., Patchala S., Pachala S., Atkar G.B., Mahalaxmi U.S.B.K. “Multi-agent learning for UAV networks: a unified approach to trajectory control, frequency allocation and routing” International Journal of Basic and Applied Sciences, 14 (2), pp. 189 - 201, 2025. https://doi.org/10.14419/474dfq89.

[26] Kailasam N., Yalamati S., Murthy V.S.N., Venkateswara Rao P., Anil Kumar R., Jayaram Kumar K. “Optimized Task Offloading in D2D-Assisted Cloud-Edge Networks Using Hybrid Deep Reinforcement Learning”International Journal of Basic and Applied Sciences, 14 (2), pp. 591 - 602, 2025. https://doi.org/10.14419/xm2ebp25.

[27] Uusitalo, Mikko A., Mårten Ericson, Björn Richerzhagen, Elif Ustundag Soykan, Patrik Rugeland, Gerhard Fettweis, Dario Sabella et al. "Hexa-X the European 6G flagship project." In 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), pp. 580-585. IEEE, 2021. https://doi.org/10.1109/EuCNC/6GSummit51104.2021.9482430.

[28] Yakushiji, Koki, Hiroshi Fujita, Mikio Murata, Naoki Hiroi, Yuuichi Hamabe, and Fumiatsu Yakushiji. "Short-range transportation using unmanned aerial vehicles (UAVs) during disasters in Japan." Drones 4, no. 4, 2020. https://doi.org/10.3390/drones4040068.
[29] Ummiti Sreenivasulu, Shaik Fairooz, R. Anil Kumar, Sarala Patchala, R. Prakash Kumar, Adireddy Rmaesh, "Joint beamforming with RIS assisted MU-MISO systems using HR-mobilenet and ASO algorithm, Digital Signal Processing, Volume 159, 2025, 104955,ISSN 1051-2004, https://doi.org/10.1016/j.dsp.2024.104955.
[30] Satyam, A. ., Kumar, R. A. ., Patchala , S. ., Pachala, S. ., Geeta Bhimrao Atkar, & Mahalaxm, U. S. B. K. . (2025). Multi-agent learning for UAV networks: a unified approach to trajectory control, frequency allocation and routing. International Journal of Basic and Applied Sciences, 14(2), 189-201. https://doi.org/10.14419/474dfq89.
[31] Kumar B. P. Mahalaxmi‹ U. S. B. K., Nagagopiraju, V., Manda, A. K., Chandana, K., Betam, D. Suresh., Gangadhar, A., & Patchala, D. S. (2025). Deep Reinforcement Learning for Joint UAV Trajectory and Communication Design in Cache-Enabled Cellular Networks. International Journal of Basic and Applied Sciences, 14(3), 418-430. https://doi.org/10.14419/djn77m90.