EAI Endorsed Transactions
on Industrial Networks and Intelligent Systems
EAI.EU Research Article

Optimizing UAV Trajectories in Optical IRS-Aided Hybrid FSO/RF Aerial Access Networks Using DRL Technique

Cuong K. P. Nguyen, Trung-Anh Do, Thang V. Nguyen, Hien T. T. Pham, and Ngoc T. Dang

AVST Lab., Posts and Telecommunications Institute of Technology, Hanoi 100000, Vietnam

Abstract

This paper investigates hybrid free-space optics (FSO)/radio frequency aerial access networks (AANs) using a high-altitude platform (HAP) and multiple UAVs to dynamically serve terrestrial users under varying environmental conditions, such as atmospheric turbulence and cloud-induced attenuation. The optical intelligent reflecting surfaces (OIRS), mounted on the HAP, enhance the FSO signal distribution to multiple UAVs by enabling precise beam manipulation, improving link reliability, and increasing network scalability. A deep reinforcement learning (DRL)-based approach is developed to optimize UAV placement and user association in real time, maximizing end-to-end throughput while adhering to backhaul capacity constraints. The study takes into account FSO channel impairments, including path loss, turbulence-induced fading, and pointing misalignment, modeled using log-normal distributions. Numerical results demonstrate that the dynamic deployment of a multi-UAV configuration, trained under realistic cloudy conditions, significantly outperforms single-UAV and static deployment strategies, achieving higher data rates and stable user connectivity. This work highlights the potential of deploying OIRS-assisted AANs supporting multiple UAVs to realize robust and high-performance 6G networks.

Received on 01 September 2025; accepted on 05 November 2025; published on 06 November 2025

Keywords: Unmanned Aerial Vehicles (UAVs), Intelligent Reflecting Surfaces (IRS), Free Space Optics (FSO), and Deep Reinforcement Learning (DRL).

doi:10.4108/eetinis.124.10134

1. Introduction

During the past decade, significant advances have been made in 5G and 6G networks, with 6G expected to offer superior capabilities such as extended coverage to remote areas, massive connectivity for IoT devices, enhanced data rates for high-definition multimedia streaming, ultra-low latency for real-time applications, and exceptional reliability for mission-critical scenarios [1, 2]. Terrestrial networks struggle to meet these needs due to coverage and infrastructure limitations, leading to the development of non-terrestrial networks (NTN) or aerial access networks (AANs) [3, 4]. AANs refer to communication networks that provide connectivity through the use of satellites, high-altitude platforms (HAP), or other non-terrestrial methods, as opposed to traditional terrestrial networks that depend on ground-based infrastructure such as cell towers. The incorporation of HAP for data backhaul and unmanned aerial vehicles (UAVs) for last-mile access represents a promising and pragmatic AAN architecture for the future 6G era, due to its rapid and flexible deployment capabilities and robust connectivity. To meet the escalating need for data rates in next-generation wireless networks, free-space optical (FSO) technology is a viable option for HAP backhaul links, while radio frequency (RF) technology plays a role in providing connections from sky to ground users for UAV access networks. Implementing a hybrid FSO/RF-based AAN marks a significant milestone in the development of the next generation of wireless networks.

The hybrid FSO/RF-based AAN aims to create a hyper-connected world by overcoming challenges in wireless communication systems [5, 6]. Despite

*Corresponding author. Email: ngocdt@ptit.edu.vn

extensive research on techniques, signal transmission methods, and complex coding methods, traditional methods still have limitations in improving the wireless transmission environment. Misalignment problems and weather-related issues, such as cloud cover and atmospheric turbulence, significantly affect the performance of the FSO link. On the one hand, the implementation of base stations mounted with UAVs presents numerous obstacles attributable to fluctuating network conditions [7]. On the other hand, the UAV must dynamically adjust its positions in response to alterations in the spatial distribution of mobile users to maintain line-of-sight (LoS) connections with elevated data rates [8]. At the same time, it might be important to consider changing weather conditions, such as moving clouds, to keep a strong connection and avoid slowing down the network. These essential challenges need effective algorithms for the optimal dynamic positioning of the UAV. Furthermore, this limitation has spurred growing interest in novel mechanisms that enable the creation of intelligent, reconfigurable, and controllable wireless propagation environments, paving the way for more robust and efficient communication systems.

With the advent of FSO/RF-based AAN as a promising solution for B5G/6G, the design of FSO backhaul and RF access has garnered significant interest, with initial research focusing on resource allocation for terrestrial users utilizing RF-based last-mile access [9–17]. Many of these studies assumed adequate backhaul link capacity, depending on the considerable transmission power of the FSO gateway within terrestrial core networks [9–11], [14]. Wu et al. specifically developed a joint optimization strategy for positioning UAVs, distributing RF bandwidth, and initiating ground users to increase the number of users who meet the guaranteed data rate requirements [9]. The authors in [10] tackled the issue of RF access bandwidth allocation, taking into account FSO backhaul for data transmission, laser charging, and the UAV-mounted base station. In [11], the issues of user association, RF bandwidth allocation, and positioning of UAV-mounted base stations were formulated and examined to minimize average latency while meeting users' quality of service (QoS) criteria. In [12], the authors examined power allocation to enhance end-to-end network throughput performance in HAP-based last-mile connectivity. In [13], the optimal placement issue of drone-based wireless networks, including 3D location optimization, was investigated through a comprehensive understanding of the challenges and the use of AI-based solutions. The BOARD algorithm was introduced to achieve suboptimal placement of backhaul-aware UAV-mounted base stations and bandwidth allocation for RF-based access in [14]. In a similar system, the authors proposed the KMAPPO design method for UAVs to ensure efficient deployment and reduced interaction overhead for emergency communication in disaster areas [15]. In addition, the study in [16] explores a hybrid satellite-HAPS-ground network, focusing on maximizing network throughput and user-association strategies by using an iterative approach and simulations to evaluate the benefits of beamforming-empowered HAPS for connecting unconnected and connected users. Last but not least, this work in [17] introduces a deep reinforcement learning framework to optimize UAV placement in AAN architecture, addressing challenges like user mobility and dynamic backhaul and access constraints.

Artificial intelligence/machine learning technologies, particularly reinforcement learning (RL) and deep reinforcement learning (DRL), offer promising solutions for adaptive placement of UAV in dynamic networks [13]. RL agents can learn optimal movement strategies through experience and trial-and-error, while DRL can handle large state spaces and high-dimensional input data. DRL excels in scenarios where the environment is unpredictable or changes over time. Therefore, in this study, we propose using DRL as an effective algorithm for the UAV placement problem in hybrid FSO/RF HAP networks. The network conditions consistently fluctuate due to user mobility and the moving cloud. Our objective is to determine the ideal positioning of the multiple UAVs to enhance the AAN's performance, considering the limits imposed by the unstable channel of the FSO backhaul and the dynamic nature of the access links. The optical intelligent reflecting surfaces (OIRS) are mounted on the HAP to enhance the FSO signal distribution by enabling precise beam manipulation. In summary, the key contributions are shown as

1. **Integrated OIRS-Assisted FSO/RF AAN with Multi-UAV**: We propose a pioneering heterogeneous network architecture that synergistically combines ground stations, a HAP equipped with OIRS, and multiple UAVs to provide seamless connectivity to terrestrial users. Unlike conventional AAN using a single UAV, our multi-UAV approach uses OIRS to enable point-to-multipoint optical signal distribution, significantly enhancing network scalability, coverage, and reliability in challenging environments.

2. **DRL-Based UAV Trajectory Optimization**: A DRL-based algorithm is introduced to optimize the trajectories of multiple UAVs and their user associations in real time. By incorporating environmental dynamics, such as cloud liquid water content (CLWC) and user mobility, the proposed algorithm maximizes end-to-end throughput while adhering to FSO backhaul capacity and

QoS constraints. The decentralized DRL framework ensures scalable and adaptive operation in non-overlapping subregions.

3. **K-Means Clustering with Geometric Constraints**: We propose a modified K-means clustering algorithm with hard geometric constraints to partition the service area into distinct subregions, each assigned to a UAV. This deterministic clustering approach guarantees feasible UAV positions and eliminates boundary ambiguity, thereby enhancing the efficiency of user association and resource allocation.

4. **Comprehensive Performance Evaluation**: Extensive simulations are conducted to evaluate the proposed framework under realistic conditions, including heterogeneous cloud scenarios and user mobility modeled via the Gauss-Markov model. The results demonstrate that the multi-UAV system, trained in cloudy environments, significantly outperforms single-UAV and static deployment strategies in terms of data rate, user connectivity, and robustness, highlighting the practical applicability of the proposed solution in 6G networks.

These contributions collectively advance the state-of-the-art in UAV-assisted wireless communications by introducing a robust, scalable, and adaptive framework that leverages OIRS and DRL to address the complex challenges of 6G connectivity. The proposed approach not only enhances network performance but also paves the way for future research on intelligent and reconfigurable wireless systems.

For the convenience of the readers, the remainder of the paper is constructed as follows. Section II first describes the system and channel model. Section III then presents the proposed DRL-based UAV position optimization. Section IV proves the effectiveness of our proposal. Finally, the key points of the study are summarized in Section V.

2. System and Channel Models

2.1. System Description

The proposed AAN architecture, which is based on hybrid FSO/RF technology to provide reliable access to terrestrial mobile users, is shown in Fig. 1. This network consists of an HAP outfitted with an optical IRS and multiple UAVs to facilitate uninterrupted high-capacity data transfer. The HAP, positioned at a significant height, functions as a central node that creates FSO backhaul connections to transmit high-speed optical signals from the core network to UAVs. The OIRS increases the FSO signal distribution by precisely modifying the beam, facilitating point-to-multipoint communication that promotes link dependability and

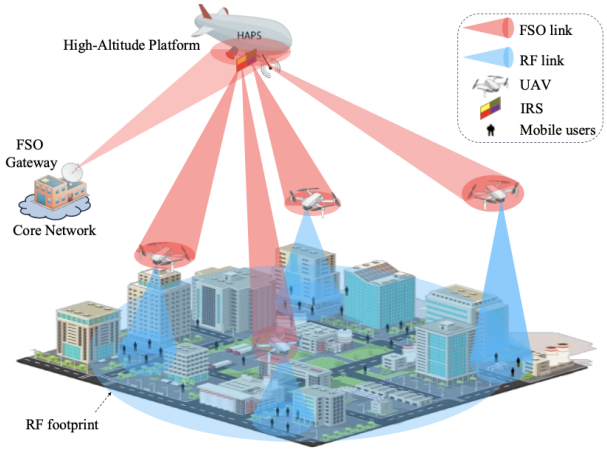

Figure 1. OIRS-assisted hybrid FSO/RF AAN based on an HAP and multiple UAVs.

network scalability. Both HAP and UAVs can operate as mobile base stations, providing RF-based access links to terrestrial users, ensuring adaptable last-mile connectivity. A portion of the mobile users are served by the UAV at high date rate, while others are supported by the HAP at a reduced data rate due to increased loss.

This hybrid design utilizes the large bandwidth of FSO links for backhaul and the adaptability of RF links for user access, mitigating the shortcomings of conventional terrestrial networks, including coverage deficiencies in remote regions. The system is engineered to accommodate dynamic user mobility and fluctuating environmental circumstances, including cloud cover and atmospheric turbulence, which may affect FSO connection performance. The deployment of multiple UAVs facilitates spatial diversity, resulting in improved load balancing and superior data rates relative to single-UAV configurations. The OIRS-assisted HAP enables adaptive signal relaying, addressing issues such as pointing misalignment and air attenuation. This comprehensive strategy guarantees resilient, high-performance connections, consistent with the objectives of ultra-low latency, extensive device connectivity, and exceptional reliability for mission-critical applications. The system functions in discrete time intervals, wherein the environmental circumstances for RF links are rather stable, whereas FSO links consider dynamic variables such as cloud liquid water content (CLWC). The multi-UAV arrangement adapts to user positions, preserving line-of-sight (LoS) connections to optimize throughput. This architecture signifies substantial progress in the development of a hyperconnected and scalable network for next-generation wireless communications.

The deployment of the hybrid FSO/RF AAN employs a DRL-based approach to optimize UAV trajectories

and user associations in real time, ensuring efficient network performance under dynamic conditions. Each UAV operates as an independent agent, trained using a decentralized DRL framework to navigate its assigned non-overlapping subregion, defined by a modified K-means clustering algorithm with geometric constraints. This clustering partitions the service area into distinct square regions, each managed by a single UAV, ensuring feasible positioning and eliminating boundary overlaps for efficient user association. The DRL algorithm accounts for environmental dynamics, such as cloud-induced attenuation and user mobility modeled via the Gauss-Markov model, to maximize end-to-end throughput while adhering to FSO backhaul capacity and quality-of-service (QoS) constraints.

The training environment incorporates realistic scenarios, including heterogeneous cloud conditions, to enhance the system's adaptability to atmospheric challenges such as turbulence and misalignment. Each UAV's DRL agent learns to adjust its position within its subregion based on user locations and channel conditions, optimizing the performance of RF access link. The decentralized approach ensures scalability, allowing the system to handle varying numbers of UAVs and users without significant computational overhead. The algorithm balances exploration and exploitation to adapt to real-time changes, such as moving clouds or shifting user distributions, ensuring stable connectivity. Using OIRS for the precise FSO signal distribution, the system enhances backhaul reliability, enabling UAVs to serve users effectively even in adverse conditions. Extensive simulations demonstrate that the multi-UAV, DRL-based approach outperforms static or single-UAV strategies, achieving higher data rates and robust user connectivity. The integration of K-means clustering ensures deterministic user assignments, further improving resource allocation efficiency. This implementation highlights the potential of intelligent, adaptive algorithms in realizing scalable and resilient 6G networks.

2.2. Channel Models

Ground-to-HAP FSO links. In the ground-to-HAP FSO link, the transmission of optical signals is primarily affected by three main factors: path loss, turbulence-induced fading, and pointing misalignment. Consequently, the overall channel coefficient h_{GH} can be expressed as

$$h_{GH} = h_{a_{GH}} h_{p_{GH}} h_{l_{GH}}, \tag{1}$$

where $h_{l_{GH}}$ represents the path loss component, which is a function of both transmission distance and prevailing weather conditions. The term $h_{p_{GH}}$ captures power loss due to pointing misalignment on the transceiver and is modeled as a random variable. Lastly, $h_{a_{GH}}$ is another random variable that describes the fluctuations in optical intensity caused by atmospheric turbulence.

As an optical signal propagates through the free space environment, it undergoes attenuation due to atmospheric absorption and scattering. The corresponding channel loss coefficient follows the Beer-Lambert law and is formulated as [18]

$$h_{l_{GH}} = \exp(-\alpha_{l_{GH}} L_{GH}), \tag{2}$$

where L_{GH} denotes the propagation distance over which the attenuation occurs and $\alpha_{l_{GH}}$ represents the attenuation coefficient, which varies based on the atmospheric conditions.

Beyond path loss, atmospheric turbulence introduces random fluctuations in the received optical signal power, leading to signal fading at the receiver. To facilitate performance analysis, various statistical models have been proposed to characterize this turbulence-induced fading. For the vertical link from ground to HAP, the log-normal distribution is commonly used due to its mathematical simplicity and precision in the weak-to-moderate regime [19]. Assuming that the fading power is normalized to unity, the probability density function (PDF) of $h_{a_{GH}}$ can be given by [18]

$$f_{h_{a_{GH}}}(h_{a_{GH}}) = \frac{1}{h_{a_{GH}} \sigma_s \sqrt{2}} \exp\left[\frac{-(\ln(h_{a_{GH}}) + \frac{\sigma_s^2}{2})^2}{2\sigma_s^2}\right], \tag{3}$$

where σ_s^2 denotes the log-intensity variance, defined as

$$\sigma_s^2 = \exp\left[\frac{0.49\sigma_R^2}{\left(1 + 0.18 z_L^2 + 0.56\sigma_R^{12/5}\right)^{7/6}} + \frac{0.51\sigma_R^2}{\left(1 + 0.9 z_L^2 + 0.62\sigma_R^{5/6}\right)^{5/6}}\right] - 1, \tag{4}$$

where σ_R^2 is the Rytov variance, which can be determined as in [20] and $z_{L_{GH}} = \sqrt{\frac{ka^2}{L_{GH}}}$.

The ergodic capacity of the FSO backhaul link, which represents the average achievable data rate, is formulated as follows

$$C_{GH}^{\text{FSO}} = \int_0^\infty B_{GH}^{\text{FSO}} \log_2(1 + \gamma_{GH}) f_\gamma^{LN}(\gamma_{GH}) d\gamma_{GH}, \tag{5}$$

where B_{GH}^{FSO} denotes the FSO bandwidth, γ_{GH} represents the instantaneous signal-to-noise ratio (SNR), and $f_{\gamma_{GH}}^{LN}(\gamma_{GH})$ is the PDF of γ_{GH} by applying the transforming variable h_{GH} with $\gamma_{GH} = \frac{(\Re \mathfrak{e} P_{FSO} h_{GH})^2}{\sigma_n^2}$, where $\Re \mathfrak{e}$ is the responsivity of the photodetector [21].

HAP-to-UAV FSO links. For the HAP-to-UAV FSO link, we consider three key impairments on the

links. including (i) cloud attenuation, (ii) atmospheric turbulence, and (iii) pointing misalignment. The combined channel fading coefficient is shown as $h_{HU} = h_{c_{HU}} h_{a_{HU}} h_{p_{HU}}$.

Cloud Attenuation: The liquid water particles in clouds induce the scattering phenomena of optical beam propagation, resulting in diminished visibility and considerable reduction of the received signal's power. This analysis focuses on low cloud types (1-3 km), such as stratus clouds and cumulus clouds, which pose significant challenges for FSO communications [22]. The attenuation caused by clouds can be articulated as [23]

$$h_{c_{HU}} = \exp[-\sigma_c H_{cl} \sec(\xi)], \tag{6}$$

where σ_c is determined by the cloud liquid water content (CLWC) and visibility V, H_{cl} is the cloud thickness, and ξ is the zenith angle of the HAP.

Atmospheric Turbulence: In an optical downlink system from the HAP, the pointing inaccuracy caused by beam wandering is typically disregarded [19]. The explanation is that the optical beam diameter exceeds that of the turbulence pocket. In the FSO uplink, where the transmitted optical beam is smaller than the turbulence eddy size, the impact of beam wander is significant. Thus, this work employs the established log-normal distribution to model beam scintillation, and the PDF of turbulence-induced fading, denoted as $h_{a_{HU}}$, can be given as

$$f_{h_{a_{HU}}}(h_{a_{HU}}) = \frac{1}{h_{a_{HU}} \sigma_R \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln h_{a_{HU}} + \sigma_R^2/2}{\sigma_R}\right)^2\right], \tag{7}$$

where σ_R^2 represents the Rytov variance [24]

Pointing Misalignment: We presume that the HAP remains constant at high altitudes, meaning that the pointing misalignment is only attributable to the UAV's hovering, influenced by wind conditions [25]. These differences lead to a repositioning of the UAV's aperture. While the UAV is in motion, its beginning position may not be at the center of the Gaussian beam footprint. Hence, the PDF of pointing misalignment, denoted as $h_{p_{HU}}$, is given by

$$f_{h_{p_{HU}}}(h_{p_{HU}}) = \frac{\phi_m^2}{h_{p_{HU}}} \left(\frac{h_{p_{HU}}}{A_m}\right)^{\phi_m^2}, \quad 0 \le h_{p_{HU}} \le A_m, \tag{8}$$

where A_m and ϕ_m are system-dependent parameters that can be found in [26]. The PDF of SNR can be written as [18]

$$f_{\gamma_{HU}}(\gamma_{HU}) = \frac{\varphi_m^2 \gamma_{HU}^{\varphi_m^2/2-1}}{4\left(A_m h_{c_{HU}} \sqrt{B_0 \bar{\gamma}_{HU}}\right)^{\varphi_m^2}} \exp\left[0.5\sigma_R^2 \varphi_m^2 (1+\varphi_m^2)\right] \operatorname{erfc}\left[\frac{\log\left(\frac{1}{A_m h_{c_{HU}}} \sqrt{\frac{\gamma_{HU}}{B_0 \bar{\gamma}_{HU}}}\right) + \mu}{\sqrt{2}\sigma_R}\right], \tag{9}$$

where μ can be defined as in [17]. The ergodic capacity of the FSO link at stage 2 can be derived as [27]

$$C_{HU}^{\text{FSO}} = \int_0^\infty B_{HU}^{\text{FSO}} \log_2(1+\gamma_{HU}) f_{\gamma_{HU}}(\gamma_{HU}) d\gamma_{HU}, \tag{10}$$

where B_{HU}^{FSO} denotes the bandwidth allocated to the FSO link for each UAV. γ_{HU} represents the instantaneous SNR, and $f_{\gamma_{HU}}^{LN}(\gamma_{HU})$ is the PDF of γ_{HU} by applying the transforming variable h_{GH} with $\gamma_{HU} = \frac{(\Re c P_{FSO}^{irs} h_{HU})^2}{\sigma_n^2}$ [21]. It is worth noting that the optical power, P_{FSO}^{irs}, is allocated to each UAV corresponding to the reflected power at each IRS element and can be given as

$$P_{FSO}^{irs} = P_{FSO} S_{irs}, \tag{11}$$

where S_{irs} is the allocated power coefficient, which is governed by the area of the IRS element [28].

RF Access Links. The air-to-ground communication links are typically modeled as probabilistic line-of-sight (LoS) channels. At any moment, a user is exclusively linked to either a UAV or a HAP. We denote the aerial platform serving user i at time t as $a_i \in \{H, U\}$, where "H" and "U" correspond to the HAP and UAV, respectively. The path loss models for LoS and non-line-of-sight (NLoS) channels in dB are given as follows [17]

$$\eta_i^{\text{LoS}}(t) = 20\log\left(\frac{4\pi f_c d_i(t)}{c}\right) + \xi^{\text{LoS}}, \tag{12}$$

$$\eta_i^{\text{NLoS}}(t) = 20\log\left(\frac{4\pi f_c d_i(t)}{c}\right) + \xi^{\text{NLoS}}, \tag{13}$$

where f_c represents the carrier frequency, while ξ^{LoS} and ξ^{NLoS} denote the average additional path loss for LoS and NLoS connections, respectively. The term $d_i(t)$ signifies the separation distance between the aerial platform and user i, which is computed as

$$d_i(t) = \sqrt{l_i(t)^2 + z_{a_i}(t)^2}, \tag{14}$$

where $l_i(t) = \sqrt{(x_{a_i}(t) - x_i(t))^2 + (y_{a_i}(t) - y_i(t))^2}$ represents the horizontal distance at time t.

The probability of a LoS connection between the user i and its serving aerial platform a_i is given by

$$P_i^{\mathrm{LoS}}(t) = \frac{1}{1 + a\exp(-b(\theta_i(t) - a))}, \tag{15}$$

where a and b are environmental-dependent parameters and $\theta_i(t) = \arctan(z_{a_i}(t)/l_i(t))$ denotes the elevation angle.

The expected path loss experienced by the user i is given by:

$$\bar{\eta}_i(t) = P_i^{\mathrm{LoS}}(t)\eta_i^{\mathrm{LoS}}(t) + (1 - P_i^{\mathrm{LoS}}(t))\eta_i^{\mathrm{NLoS}}(t). \tag{16}$$

The data rate of user i at time t is estimated using the Shannon-Hartley theorem

$$R_i(t) = \frac{B^{\mathrm{RF}}}{N}\log_2\left(1 + \frac{P_{t,a_i}^{\mathrm{RF}}10^{-\bar{\eta}_i(t)/10}}{N_0}\right), \tag{17}$$

where B^{RF} is the total allocated RF bandwidth, P_{t,a_i}^{RF} represents the transmit power of the aerial platform, and N_0 is the noise power in the downlink channel. The bandwidth is assumed to be uniformly distributed among all ground users. From equations (2), (6), and (12), the FSO capacity can be derived as follows

$$R_i(t) = \frac{B_{HU}^{\mathrm{RF}}}{N}\log_2\left(1 + \frac{P_{t,a_i}^{\mathrm{RF}}10^{-\bar{\eta}_i(t)/10}}{N_0}\right), \tag{18}$$

User Movement: The movement of users can be effectively modeled using the Gauss-Markov mobility model [29]. At time slot t, the speed and direction of user i are determined by their values from the previous time slot $(t-1)$, expressed as

$$s_i(t) = \alpha s_i(t-1) + (1-\alpha)\bar{s} + \sqrt{1-\alpha^2}s_{n,i}(t-1), \tag{19}$$
$$d_i(t) = \alpha d_i(t-1) + (1-\alpha)\bar{d} + \sqrt{1-\alpha^2}d_{n,i}(t-1), \tag{20}$$

Here, $s_i(t)$ and $d_i(t)$ represent the speed and direction of user i at time step t. The parameters α, $\bar{s}$, and $\bar{d}$ define the memory level, the average speed, and the mean direction, respectively. The terms $s_{n,i}(t-1)$ and $d_{n,i}(t-1)$ are Gaussian-distributed random variables [17].

The user's position updates at each time step according to the following equations

$$x_i(t) = x_i(t-1) + s_i(t-1)\cos d_i(t-1), \tag{21}$$
$$y_i(t) = y_i(t-1) + s_i(t-1)\sin d_i(t-1), \tag{22}$$

where $x_i(t)$ and $y_i(t)$ correspond to the coordinates of user i at time slot t.

3. DRL-based UAV Placement

3.1. UAV Placement Model

Figure 2 illustrates the operational mechanism of the deep Q-learning (DQN) algorithm in the context of UAVs establishing FSO links with a HAP. DQN integrates traditional Q-Learning with deep neural networks, enabling UAVs to learn optimal policies in complex environments without requiring an explicit model of the communication channel. Unlike the conventional Q-table approach, which becomes impractical for large or continuous state spaces, DQN employs a neural network to approximate Q-values for all state-action pairs, thereby overcoming the limitations of classical methods.

The distance between the UAVs and the users plays a key role in the network performance. To ensure UAVs serve users within predefined square regions, we modify the K-means clustering algorithm with hard geometric constraints. Given K square regions $\{Q_1, Q_2, ..., Q_K\}$ of side length L, the algorithm proceeds as follows [30, 31]

- **Cluster Assignment** Assign each user q_u to a cluster S_k if and only if q_u lies within Q_k:

$$S_k = \{q_u \mid q_u \in Q_k\}, \quad \forall k \in \{1, 2, ..., K\}. \tag{23}$$

- **Centroid Calculation** Compute the centroid c_k for each cluster, guaranteed to lie within Q_k:

$$c_k = \begin{cases} \frac{1}{|S_k|}\sum\limits_{q_u \in S_k} q_u & \text{if } |S_k| > 0, \\ \mathrm{Center}(Q_k) & \text{otherwise,} \end{cases} \tag{24}$$

where $\mathrm{Center}(Q_k)$ is the geometric center of Q_k.

The optimization objective becomes

$$\min \sum_{k=1}^{4}\sum_{q_u \in S_k} \|q_u - c_k\|^2 \text{ subject to } S_k \subseteq Q_k,\ c_k \in Q_k. \tag{25}$$

Key Properties:

- **Deterministic Clustering**: Users are strictly assigned to squares, eliminating boundary ambiguity.
- **Feasible UAV Positions**: All centroids c_k lie within their designated regions Q_k.

3.2. Problem Formulation

To extend the single-UAV placement problem, we consider a scenario with M UAVs, where each UAV independently serves users in a distinct, non-overlapping subregion of the overall service area. The

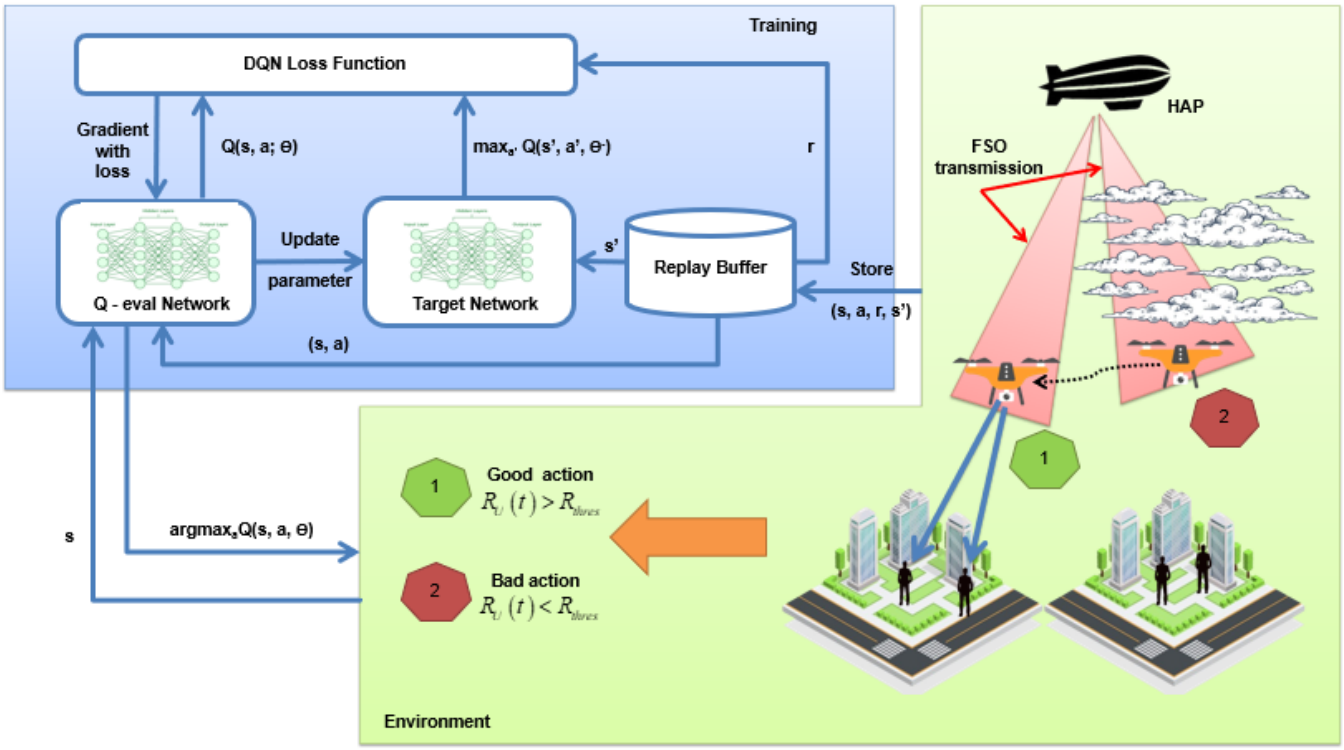

Figure 2. Diagram of the deep Q-learning algorithm operation for UAV trajectory optimization.

goal remains to maximize the cumulative data rate of all users served by all UAVs over a total time horizon T.

Let $\mathcal{U}_t^{(m)}$ denote the set of users served by the UAV $m \in \{1, 2, ..., M\}$ at time step t, and $(x_U^{(m)}(t), y_U^{(m)}(t))$ be the position of the UAV m within its designated area $\mathcal{A}^{(m)}$. The optimization problem for each UAV is defined independently as

$$\max_{\{x_U^{(m)}(t), y_U^{(m)}(t)\}} \quad \sum_{t=1}^{T} \sum_{i \in \mathcal{U}_t^{(m)}} R_i^{(m)}(t) \tag{26a}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{U}_t^{(m)}} \mathbb{I}(c_i^{\text{FSO},(m)}(t)), \quad \forall t \tag{26b}$$

$$\sum_{i \in \mathcal{U}_t^{(m)}} R_i^{(m)}(t) \leq C_{\text{FSO}}^{(m)}(t), \quad \forall t \tag{26c}$$

$$(x_U^{(m)}(t), y_U^{(m)}(t)) \in \mathcal{A}^{(m)}, \quad \forall t \tag{26d}$$

where Eq. (26*a*) is to maximize the total data rate that UAV m delivers to users $i \in U_t^{(m)}$ over the entire time period T , where $R_i^{(m)}$ is the user data rate at time t. Constraint (26*b*) uses a directive function $I(c_{FSO,i}^{(m)}(t))$ to ensure that only users with a feasible FSO backhaul connection (requested capacity, $c_{FSO,i}^{(m)}$, does not exceed available capacity) are served. Other constraints will be explained in the next section.

3.3. Constraints

To ensure the practical feasibility and optimal performance of the proposed DRL-based UAV placement algorithm in the OIRS-assisted AAN, several key constraints are imposed. These constraints reflect the physical, operational, and environmental limitations of the multi-UAV architecture and are critical for achieving reliable and efficient network operation.

1. **FSO Backhaul Capacity Constraint** (26c): The total data rate served by each UAV m to its associated users $\mathcal{U}_t^{(m)}$ at time t, denoted as $\sum_{i \in \mathcal{U}_t^{(m)}} R_i^{(m)}(t)$, must not exceed the instantaneous FSO backhaul capacity $C_{\text{FSO}}^{(m)}(t)$. This constraint accounts for the limitations imposed by atmospheric conditions such as clouds, fading, and misalignment. The FSO capacity $C_{\text{FSO}}^{(m)}(t)$ is derived from the ergodic capacity of the FSO link and is influenced by the channel gain $h = h_a h_c h_p$.
2. **UAV Positioning Constraints** (26d): Each UAV $m \in \{1, 2, ..., M\}$ must operate within its designated subregion $\mathcal{A}^{(m)}$, which is a predefined square area of side length L. This ensures non-overlapping UAV operations and prevents inter-UAV interference. Each UAV is trained and operated independently using a separate DRL agent. The environment for each agent includes only

the users and atmospheric conditions within its assigned region $\mathcal{A}^{(m)}$. This decentralized formulation allows scalable training while ensuring non-overlapping UAV operation. Additionally, the altitude of each UAV is constrained to a fixed range to maintain line-of-sight connectivity with the HAP and ground users while adhering to regulatory altitude limits, typically set between 100 m and 500 m.

3. **RF Bandwidth Allocation Constraint**: The total RF bandwidth B^{RF} is uniformly allocated among the N ground users served by the UAVs and HAP. For each user i, the allocated bandwidth is $\frac{B^{\mathrm{RF}}}{N}$, ensuring fair resource distribution. This constraint limits the achievable data rate $R_i(t)$ for each user and is critical for maintaining equitable QoS across the network.

4. **User Association Constraint**: At each time slot t, each user i is exclusively associated with either a UAV or the HAP, denoted as $a_i \in \{H, U\}$. The greedy-based user association algorithm (Algorithm 1) ensures that only users within the UAV's coverage radius $R_{\mathrm{covered}} = 100\,\mathrm{m}$ are considered for high-speed service, subject to the available FSO backhaul capacity. Users outside this radius or when backhaul capacity is exhausted are served by the HAP via RF links, which typically offer lower throughput due to higher path loss.

5. **Environmental Constraints**: The system must operate under dynamic environmental conditions, including cloud-induced attenuation and atmospheric turbulence. The cloud liquid water content (CLWC), ranging from 0.5 mg to 7.5 mg, and its spatial variation across a 100 m × 100 m grid, impose time-varying attenuation on the FSO links, as modeled by the Beer-Lambert law. Similarly, turbulence-induced fading, characterized by log-normal or gamma-gamma distributions, affects the FSO channel reliability. These conditions are incorporated into the DRL environment as input features to ensure robust UAV trajectory planning.

6. **QoS Requirements**: To meet the stringent demands of 6G applications, the system enforces a minimum data rate threshold R_{thres} for users served by UAVs. The reward function is designed to penalize scenarios where the total data rate $R_U(t)$ falls below this threshold, ensuring that the DRL agent prioritizes QoS-aware UAV positioning and user association.

3.4. User Association—Greedy-Based Approach

When a user falls within the coverage area of a UAV, it is eligible to receive high-speed service from that UAV. Nevertheless, due to limitations in the backhaul capacity, not all users within coverage can be supported simultaneously. To enhance network performance, a selected group of users must be chosen wisely. This is achieved through a greedy strategy (i.e., Algorithm 1) that prioritizes users based on their achievable data rates—those with higher rates are given precedence. The algorithm proceeds as follows:

- **Step 1:** Identify the set of users $\mathcal{U}_{\mathrm{cov}}$ that lie within the UAV's coverage area. Initialize two disjoint sets: $\mathcal{U}_{\mathrm{U}} = \{\}$ for users to be served via UAV and $\mathcal{U}_{\mathrm{H}} = \{\}$ for users to be served directly by the HAP. Sort the users in $\mathcal{U}_{\mathrm{cov}}$ in descending order based on their data rate requirements $R_u(t)$.
- **Step 2:** For each user u in the sorted list, check whether the UAV still has sufficient FSO backhaul capacity to forward the traffic of user u to the HAP:
  - If $R_u(t) \le C^{\mathrm{FSO}}(t)$, associate user u with the UAV, add u to the set $\mathcal{U}_{\mathrm{U}}$, and update the remaining backhaul capacity:
    $$C^{\mathrm{FSO}}(t) = C^{\mathrm{FSO}}(t) - R_u(t).$$
  - Otherwise, user u is connected directly to the HAP and added to the set $\mathcal{U}_{\mathrm{H}}$.
- **Step 3:** Repeat Step 2 for all users $\mathcal{U}_{\mathrm{cov}}$ until every user has either been assigned to the UAV or directly to the HAP.

3.5. The DRL-based Algorithm

To address the formulated MDP problem, this work employs the Q-learning framework, which is a model-free approach that estimates the expected cumulative reward through iterative interaction with the environment. However, due to the high-dimensional state space in the considered system, maintaining a conventional Q-table becomes infeasible in terms of memory and computational requirements. To mitigate this limitation, the Deep Q-Network (DQN) [32] is adopted, where a deep neural network (DNN) approximates the Q-function to effectively handle large-scale state–action spaces and capture their nonlinear dependencies.

In the DQN framework, the network parameters are trained by minimizing the following loss function

$$L(\theta) = \left(y_t - Q(s_t, a_t; \theta)\right)^2, \tag{27}$$

where θ represents the weight parameters of the Q-network and $Q(s_t, a_t; \theta)$ denotes the estimated Q-value

Algorithm 1 Greedy User Association Algorithm with HAP Fallback

Input: User set within UAV coverage $\mathcal{U}_{\text{cov}}$, total FSO backhaul capacity $C^{\text{FSO}}(t)$ to HAP
Output: Associated user sets: $\mathcal{U}_{\text{U}}$ (served via UAV), $\mathcal{U}_{\text{H}}$ (served directly by HAP)
Initialize $\mathcal{U}_{\text{U}} \leftarrow \emptyset, \mathcal{U}_{\text{H}} \leftarrow \emptyset$
Sort users in $\mathcal{U}_{\text{cov}}$ in descending order of $R_u(t)$
For each user u in sorted, $\mathcal{U}_{\text{cov}}$
  If $R_u(t) \leq C^{\text{FSO}}(t)$
  $\mathcal{U}_{\text{U}} \leftarrow \mathcal{U}_{\text{U}} \cup \{u\}$
  Establish FSO link from UAV to HAP for the user. u
  $C^{\text{FSO}}(t) \leftarrow C^{\text{FSO}}(t) - R_u(t)$
  Else
  $\mathcal{U}_{\text{H}} \leftarrow \mathcal{U}_{\text{H}} \cup \{u\}$ Connect user u directly to HAP
  End
End
return $\mathcal{U}_{\text{U}}, \mathcal{U}_{\text{H}}$

Algorithm 2 DQN-based UAV trajectory

Initialize the replay buffer D with capacity C
Initialize the Q network with random weights θ
Initialize the target $\hat{Q}$ network with weights $\hat{\theta} = \theta$
For each episode
Initialize the network environment and receive the initial state s_0
  For each time slot t
  Obtain action a_t according to the ϵ greedy policy.
  Execute a_t and perform user association algorithm
  Observe reward r_t and next state s_{t+1}
  Store transition (s_t, a_t, r_t, s_{t+1}) in replay buffer D
  Sample a random minibatch of transitions from D
  Obtain the target Q-value y_i according to (18)
  Perform a gradient descent step on (17) with respect to θ
  Reset $\hat{Q} = Q$ every fixed number of steps
  End
End

at time step t. The target Q-value, y_t, is computed as

$$y_t = \begin{cases} r_t, & \text{if the episode terminates at step } t+1, \\ r_t + \gamma \max_{a'} \hat{Q}(s_{t+1}, a'; \hat{\theta}), & \text{otherwise,} \end{cases} \tag{28}$$

where γ is the discount factor that controls the influence of future rewards and $\hat{\theta}$ represents the parameters of the target network $\hat{Q}$. The detailed procedure of the DRL-based UAV placement strategy is provided in Algorithm 2.

3.6. Performance Metrics

To evaluate the effectiveness of the proposed DRL-based UAV placement algorithm, we consider several key performance metrics that reflect the overall system performance and learning efficiency of the agent.

End-to-End Throughput. The primary performance metric is the *end-to-end throughput*, which measures the total data rate achieved by all mobile users (MUs) in the network. This accounts for both the radio access link and the constraints of the FSO backhaul link. The throughput is computed as the sum of the data rates of all users over a given time duration

$$\text{Total throughput} = \sum_{t=0}^{T-1} \sum_{i=1}^{N} R_i(t), \tag{29}$$

where $R_i(t)$ denotes the data rate of user i in the time slot t, and T is the total number of time slots.

Immediate Reward. To facilitate learning, the DRL agent receives a reward at each time step, which serves as a performance indicator during training. The reward function is designed to encourage both high access data rates and sufficient backhaul capacity

$$r(s_t, a_t) = \begin{cases} \alpha \sqrt{R_U(t) C_{\text{FSO}}(t)}, & \text{if } R_U(t) \geq R_{\text{thres}} \\ 1 - \alpha \sqrt{R_U(t) C_{\text{FSO}}(t)}, & \text{otherwise,} \end{cases} \tag{30}$$

where $R_U(t)$ is the total data rate served by the UAV at time t, $C_{\text{FSO}}(t)$ is the FSO backhaul capacity, α is a scaling factor, and R_{thres} is a predefined threshold.

Episode Return. The cumulative reward obtained over a training episode, known as the *episode return*, is also used as a performance metric. It reflects the overall learning progress of the DRL agent. A higher episode return indicates better decision-making in UAV positioning across time steps.

Average User Throughput. In addition to cumulative metrics, we also evaluate the average data rate per user at each time step, calculated over multiple test runs under various atmospheric turbulence and environmental attenuation conditions. This metric facilitates a comparative analysis between the proposed method and baseline scenarios, such as

- Fixed UAV located at the center
- Services provided by HAP
- Training under cloudy conditions
- Training without considering cloud effects

To highlight the effectiveness of DQN-based UAV trajectory, the coordinated trajectories of multiple UAVs reflect the effectiveness of the modified K-means clustering algorithm with geometric constraints,

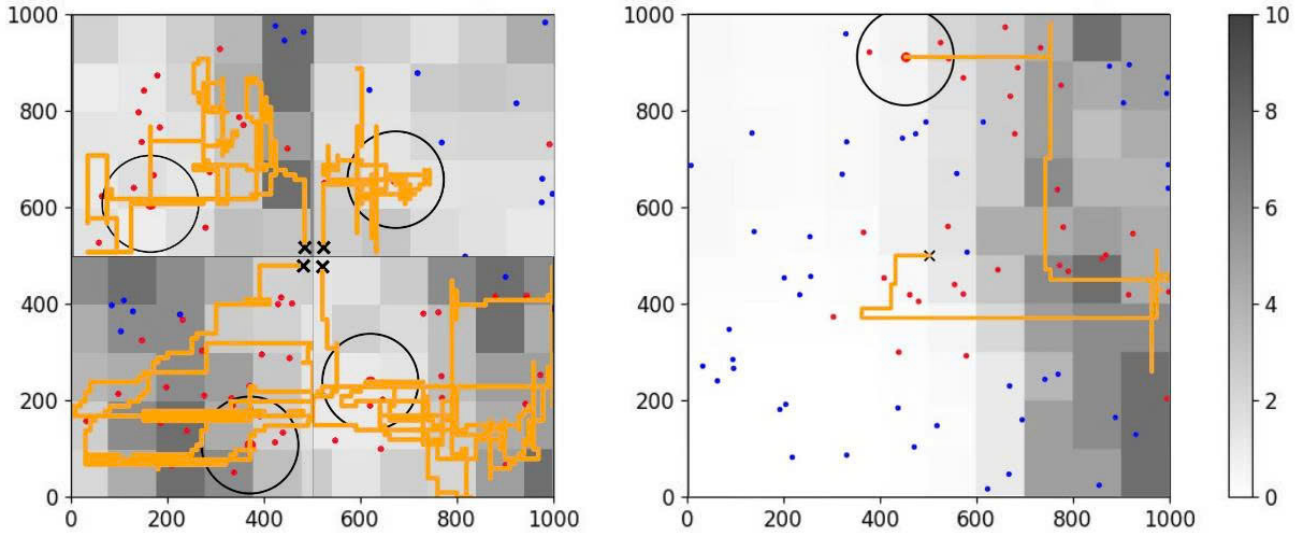

Figure 3. UAV flight trajectories and user coverage in multi-UAV and single-UAV scenarios

ensuring non-overlapping subregions and efficient user association, as shown in Algorithm 2. The limited range of a single UAV highlights its inadequacy for large-scale coverage, leaving many users unserved. As shown in Fig. 3, the finding directly supports the third contribution by illustrating how K-means clustering with geometric constraints optimizes UAV positioning and user association.

The left subfigure depicts the flight trajectories and coverage regions in the multi-UAV scenario, where four UAVs adaptively navigate within non-overlapping subregions (the black circles represent the coverage areas of individual UAVs). Each UAV is responsible for serving users within its assigned region—red dots indicate served users, while blue dots represent unserved users—thereby minimizing coverage gaps, reducing the number of unserved users, and avoiding coverage conflicts among UAVs. In contrast, the right subfigure illustrates the single-UAV scenario, where a single UAV attempts to cover the entire operational area but fails to meet user demands effectively due to its limited coverage range. The yellow lines represent the UAV flight trajectories, demonstrating that UAVs in the multi-UAV scenario achieve more efficient mobility and better coverage optimization compared to the single-UAV case.

The ability of multi-UAV-based AAN to minimize coverage gaps demonstrates its suitability for 6G networks requiring massive connectivity. Visual evidence reinforces the need for multiple UAVs to achieve robust and scalable AANs.

These constraints collectively ensure that the proposed system operates within realistic operational boundaries while maximizing end-to-end throughput and maintaining QoS. The DRL-based approach leverages these constraints to learn adaptive UAV trajectories that dynamically respond to user mobility, environmental variations, and network resource limitations, thereby enhancing the scalability and reliability of the OIRS-assisted HAP communication framework.

Table 1. System Parameters

Parameter (Symbol)	Value
FSO bandwidth (B_{FSO})	3×10^9 Hz
Standard deviation (σ_n)	10^{-7} A/Hz
FSO transmit power (P_{FSO})	12 dBm
Wavelength (λ)	1550 nm
Optical wave number (K_{wave})	$2\pi/\lambda$
Transmitter aperture diameter (D_t)	0.08 m
Turbulent strength (C_n^2)	10^{-13} $m^{-2/3}$
Beam divergence angle (θ)	10^{-3} rad
UAV coverage radius (R_{covered})	100 m
Bandwidth per user (bandw_{GU})	350 MHz
Noise power spectral density (N_0)	−104 dBm/Hz
RF transmit power of UAV ($P_{t,\text{UAV}}$)	25 dBm
RF transmit power of HAP ($P_{t,\text{HAP}}$)	35 dBm
Active zone	1 x 1 km

4. Numerical Results

The proposed DRL-based UAV placement algorithm, incorporating an IRS mounted on the UAV, is evaluated in simulations for single-UAV and multi-UAV deployment scenarios. The simulation area spans 1 km × 1 km and includes 75 users with a Gaussian distribution with 20% of them exhibiting mobility

patterns, while the remaining users remain stationary. A high-altitude platform (HAP) is located at an altitude of 20 km. We consider the case of four UAVs whose initial positions are randomly initialized during the training phase. Cloud conditions are considered, with CLWC levels varying from 0.5 mg to 7.5 mg. The cloud projection is modeled as a grid, with each 100 m × 100 m cell assigned a specific CLWC value. The average CLWC value within each grid cell is used as an input feature for the DQN algorithm. Otherwise, the detailed parameters are shown in Table 1.

Backhaul link: The optical backhaul employs a transmitted power of $P_t^{\mathrm{FSO}} = 12$ dBm with a bandwidth $B^{\mathrm{FSO}} = 3$ GHz. The divergence angle of the HAP's optical beam is set to $\theta_t = 1$ mrad, while the pointing jitter is $\theta_j = 50$ μrad. The optical wavelength is $\lambda = 1.55$ μm, and the UAV's aperture diameter is $D_r = 8$ cm. The noise spectral density is $\sigma_n = 10^{-7}$ A/Hz, with a turbulence coefficient of $C_n^2 = 10^{-13}$ m$^{-2/3}$. The cloud droplet concentration is $N_c = 250$ cm^{-3}, the atmospheric altitude is $H_a = 20$ km, the cloud thickness is $H_c = 2$ km, and the wind velocity is $v_{\mathrm{wind}} = 21$ m/s.

Access link: For the RF access connection, the transmit powers are $P_{t,H}^{\mathrm{RF}} = 35$ dBm and $P_{t,U}^{\mathrm{RF}} = 25$ dBm. The total RF bandwidth is $B^{\mathrm{RF}} = 350$ MHz, with a carrier frequency of $f_c = 2$ GHz. The environmental constants (a, b) are $(9.61, 0.16)$, the average excess path losses $(\xi_{\mathrm{LoS}}, \xi_{\mathrm{NLoS}})$ are (1 dB, 20 dB), and the noise power is $N_0 = -80.4$ dBm.

First, the training performance of a UAV control model using reinforcement learning under optimal conditions (without and with cloud cover) is demonstrated in Fig. 4. The DRL model attains a consistent reward plateau of 150-160 after around 300,000 episodes, indicating a resilient control policy. Under overcast conditions, convergence occurs more slowly, stabilizing at a diminished reward of approximately 15 with more variance due to environmental uncertainty. This swift convergence under optimal conditions demonstrates the efficacy of the DRL method in acquiring optimal UAV trajectories when environmental variables are limited. However, the diminished convergence rate and reduced reward in overcast conditions indicate the complexity caused by cloud-related attenuation and turbulence, affecting the model's generalization capability. The increased variance in overcast conditions, underscoring its adaptability under realistic scenarios, underscores the importance of training under realistic scenarios for 6G networks operating in dynamic environments.

Fig. 5 shows the total data rates of users served by a single UAV over time, though specific details are not provided due to truncation. Here, we consider four situations, including training the UAV with moving clouds (blue line), without moving clouds (orange line), fixing UAVs at the center of the areas (green line), and services provided by HAP, i.e. no UAV implementation (red line). This result illustrates the cumulative data rates of MUs for a duration of 1000 seconds (one test episode). The testing environment is established arbitrarily, and the UAV has no previous knowledge of this environment. Our trained agent consistently demonstrates superior performance and maintains a reasonably high data rate over time. This indicates that the taught agent can maneuver well, resulting in substantial disparities with other instances.

Next, the comparable data rates of multi-UAV-based AAN, despite the reduced bandwidth per UAV, highlight the efficiency of spatial diversity and the distribution of the FSO signal due to OIRS, as shown in Fig. 6. The multi-UAV-based AAN (see figure on the left side) offers a more extensive and equal data rate distribution than the single-UAV-based AAN (see figure on the right side), even though each UAV utilizes one-quarter of the bandwidth. The single-UAV-base AAN shows high data rates near the UAV but significant declines at farther distances, leaving many areas underserved. This result strongly supports the integrated OIRS-assisted multi-UAV architecture by showcasing its ability to enhance network scalability and throughput. The OIRS's role in precise beam manipulation enables efficient FSO backhaul, allowing multiple UAVs to serve users effectively. The comparison underscores the limitations of single-UAV systems in 6G networks, where widespread connectivity is crucial.

For connectivity and resource allocation efficiency in our considered system, Fig. 7 and Fig. 8 show that the higher and more stable user count of multi-UAV-based AAN reflects its ability to effectively handle dynamic user mobility and environmental challenges. The single-UAV-based AAN's instability highlights its limited coverage and adaptability. The multi-UAV-based AAN serves 3–6 users per time step with stable connectivity, while the single-UAV system fluctuates between 0–2 users, indicating inconsistent performance. This result demonstrates the multi-UAV system's superior connectivity and the effectiveness of K-means clustering in user association. The decentralized DRL approach enables adaptive operation, reducing service interruptions. The single-UAV's performance drops emphasize the scalability benefits of multiple UAVs. Real-time user tracking could further enhance user association efficiency.

Figure 9 presents the average user data rates per time step in the proposed OIRS-assisted multi-UAV-based AAN, optimized through DRL. The findings demonstrate the network capacity to provide high and stable data rates in dynamic conditions, including user mobility and cloud-related FSO impairments, thereby fulfilling QoS requirements. The multi-UAV configuration, improved through OIRS beam manipulation and

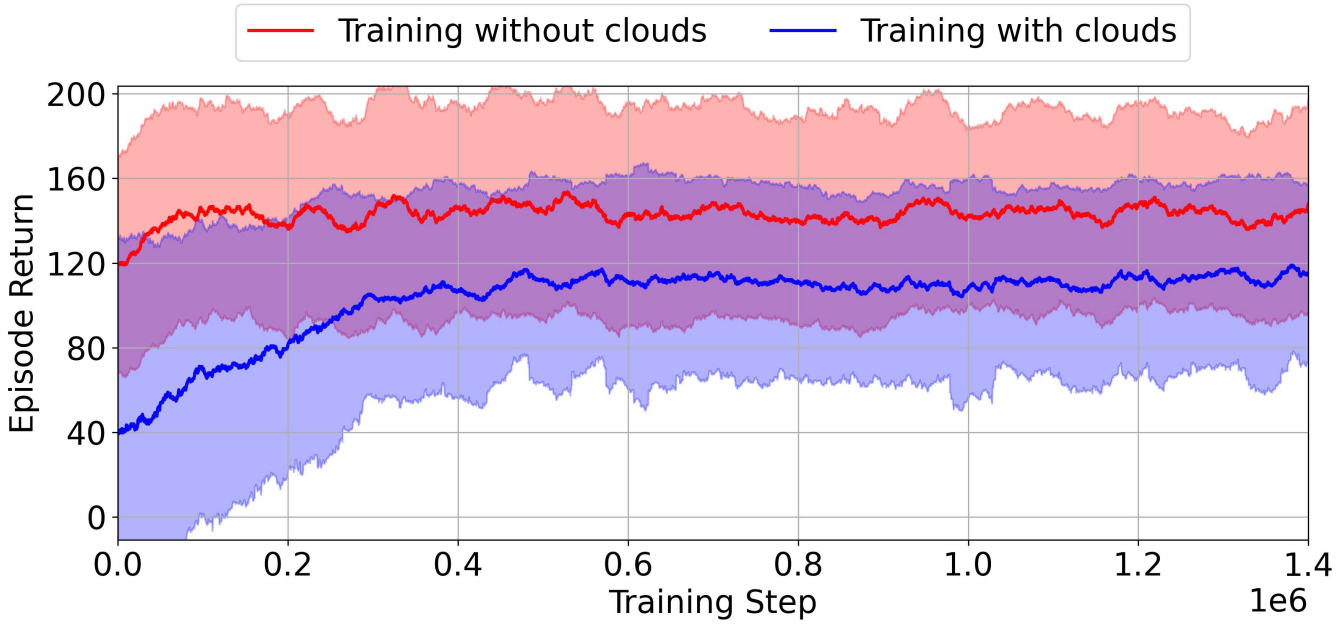

Figure 4. Episode reward versus training episodes

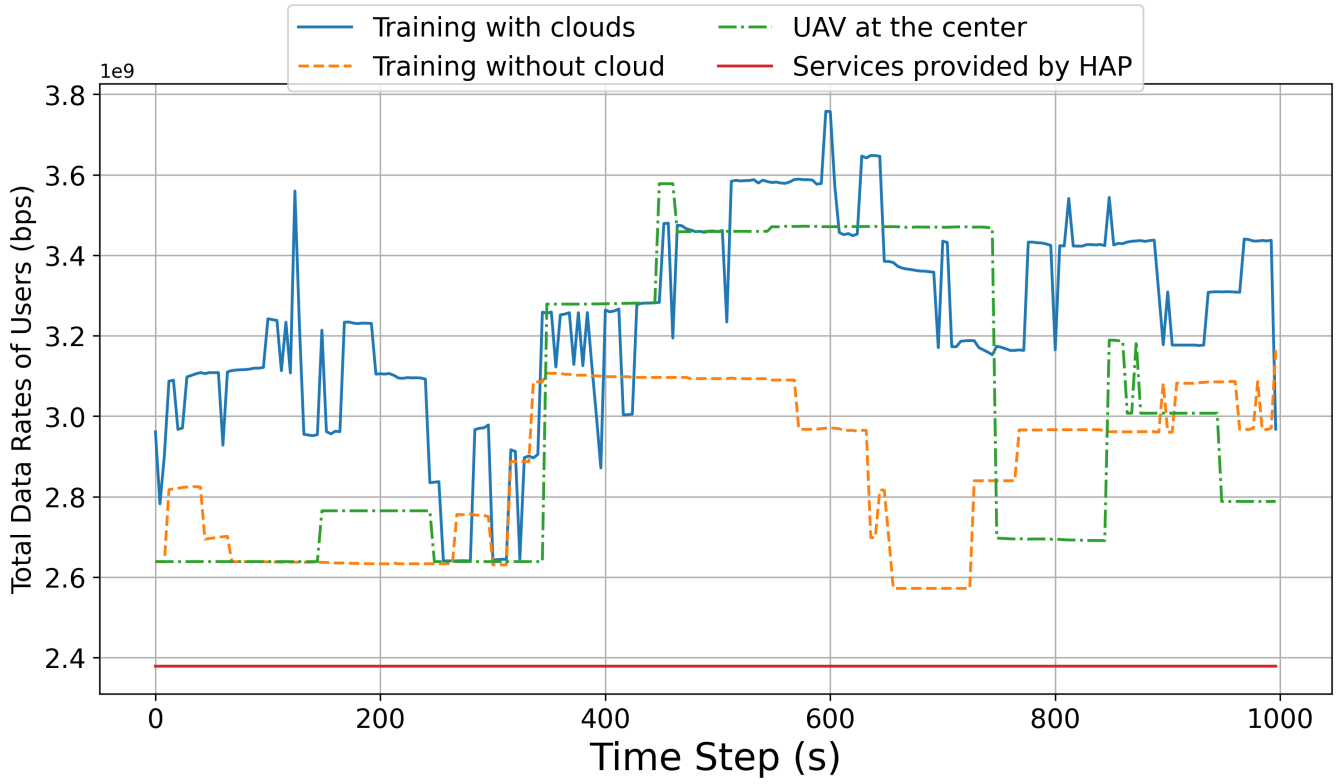

Figure 5. Time-varying total data rate achieved by users in OIRS-assisted AAN

DRL-based trajectory optimization, demonstrates superior performance compared to single-UAV and static deployments. The decentralized DRL framework, in conjunction with K-means clustering, facilitates efficient user association and resource allocation, thereby maximizing throughput while complying with FSO backhaul and QoS constraints. The OIRS addresses

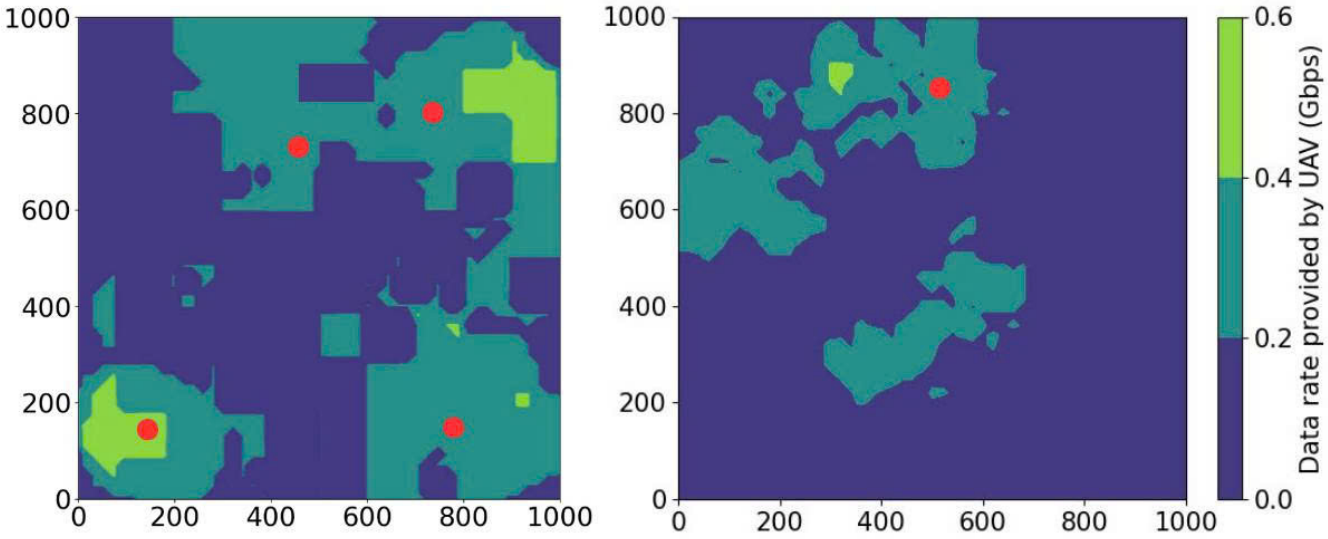

Figure 6. Data rate distribution by OIRS-assisted hybrid FSO/RF AANs using: multi-UAV (left) versus single-UAV (right).

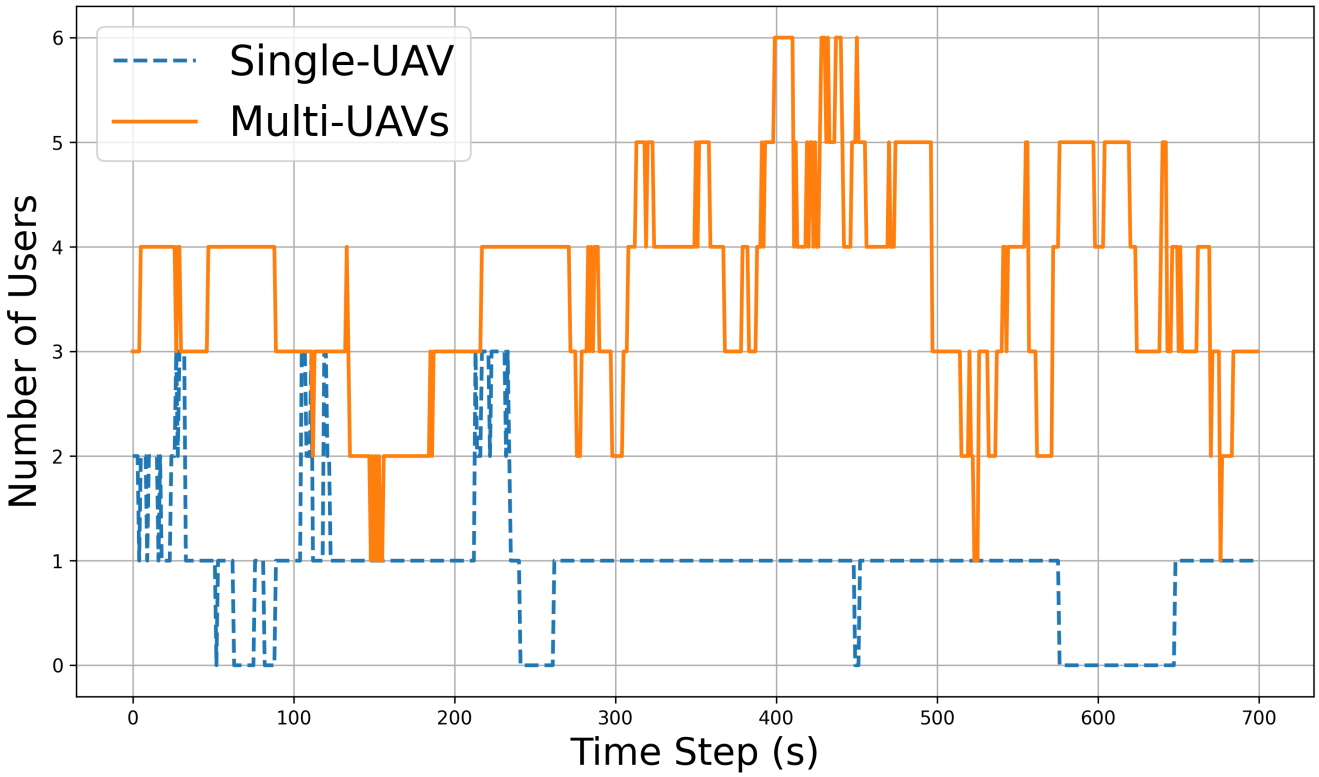

Figure 7. Temporal analysis of user data rates in a single-UAV OIRS-assisted multi-UAV AAN.

FSO link challenges such as turbulence and misalignment, enhancing both reliability and scalability. This result demonstrates the framework's robustness, as evidenced by consistent data rates that indicate its adaptability to environmental challenges. The multi-UAV approach demonstrates superior connectivity and performance when contrasted with single-UAV systems, which exhibit variable rates and restricted coverage. DRL training complexity indicates a need for future research focused on enhancing efficiency and optimizing bandwidth allocation. Finally, this outcome confirms the potential of the OIRS-assisted multi-UAV

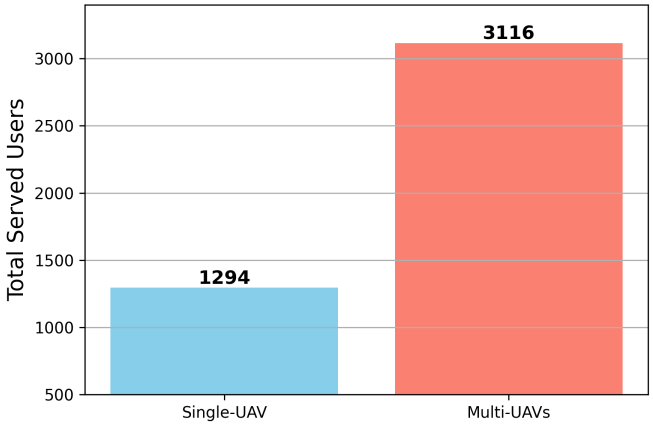

Figure 8. Comparison of user connectivity: multi-UAV versus single-UAV in dynamic network environments.

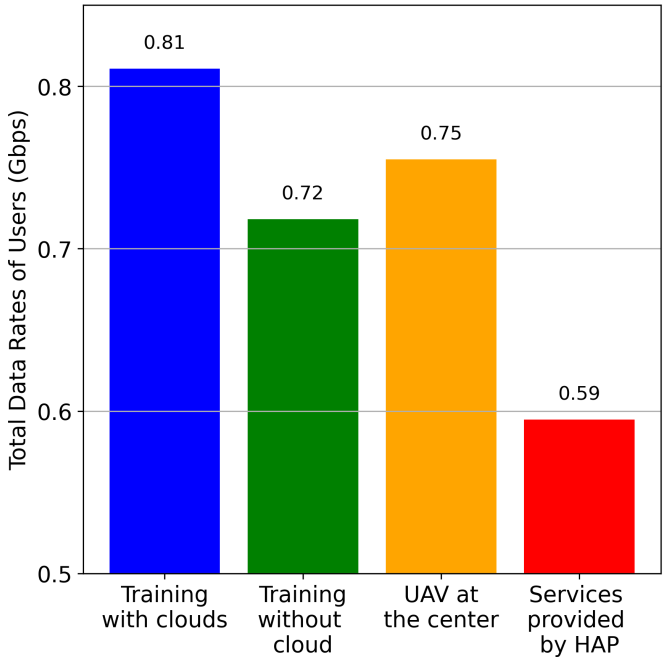

Figure 9. Average user data rates per time step in OIRS-assisted multi-UAV networks.

system for scalable, high-performance 6G networks, facilitating further investigation into intelligent NTN.

5. Conclusion

This paper presents a QoS-aware trajectory planning framework for multi-UAV coupled with OIRS in hybrid FSO/RF AAN, employing DRL to fulfill 6G connectivity requirements. The proposed approach enhances UAV trajectories to maximize throughput while complying with FSO backhaul and QoS constraints, surpassing single-UAV and static deployment solutions. Numerical findings indicate enhanced data rates, user coverage, and resilience under actual scenarios such as cloud attenuation and user mobility. The OIRS improves FSO link dependability and scalability via accurate beam manipulation. The decentralized DRL methodology guarantees flexible and scalable functionality in dynamic settings. Notwithstanding obstacles like training difficulty, the framework underscores the promise of OIRS-assisted multi-UAV AAN for 6G. Subsequent efforts must focus on improving training efficiency and optimizing bandwidth to improve performance. This research facilitates the development of intelligent, dependable AANs in next-generation wireless communications.

The proposed OIRS-assisted hybrid FSO/RF AAN utilizing multi-UAV DRL optimization enhances 6G throughput and reliability, though significant extensions are required. Future endeavors may incorporate energy-efficient UAVs, such as solar-powered High Altitude Platforms (HAPs), to promote sustainability; augment security through blockchain authentication and Free Space Optics (FSO) encryption; expand to heterogeneous Non-Terrestrial Networks (NTNs) utilizing federated Deep Reinforcement Learning (DRL) for comprehensive coverage; perform field trials in extreme weather conditions to validate models; and investigate graph neural networks for multi-agent collaboration in Internet of Things (IoT) contexts. These enhance our efforts to facilitate robust aerial networks for intelligent urban environments and emergency response.

Acknowledgement

This research is funded by Vietnam Ministry of Science and Technology under Project DT.34/25.

References

[1] Nguyen, T.V., Le, H.D. and Pham, A.T. (2023) On the design of ris–uav relay-assisted hybrid fso/rf satellite–aerial–ground integrated network. *IEEE Transactions on Aerospace and Electronic Systems* **59**(2): 757–771. doi:10.1109/TAES.2022.3189334.

[2] Zhou, D., Sheng, M., Li, J. and Han, Z. (2023) Aerospace integrated networks innovation for empowering 6g: A survey and future challenges. *IEEE Communications Surveys & Tutorials* **25**(2): 975–1019. doi:10.1109/COMST.2023.3245614.

[3] France, T.A.S., Sweden, E., France, E.F.Q., Techcom, S., SIX, T., Cinterion, T., GreenerWave *et al.* (2024) Vision on non-terrestrial networks in 6g system (or imt-2030). In *ETSI Conference on Non-Terrestrial Networks, a Native Component of 6G*: 1–5.

[4] (2024), 3rd generation partnership project (3gpp) technologies - non-terrestrial networks (ntn), [Online]. Available: https://www.3gpp.org/technologies/ntn-overview.

[5] Khoshafa, M.H., Bueno, F., Ngatched, T.M.N. and Di Renzo, M. (2025) Ris-empowered secured space-air-ground integrated networks: Opportunities and

challenges. *IEEE Communications Magazine* **63**(6): 130–136. doi:10.1109/MCOM.001.2400398.

[6] Nguyen, T.V., Le, H.D., Dang, N.T. and Pham, A.T. (2021) On the design of rate adaptation for relay-assisted satellite hybrid FSO/RF systems. *IEEE Photon. J.* : 1–1doi:10.1109/JPHOT.2021.3130720.

[7] Dabiri, M.T. and Hasna, M. (2025) A novel mrr-uav-based relay with optical network coding: A comparative study with optical irs and conventional uav relaying. *IEEE Journal on Selected Areas in Communications* **43**(5): 1607–1620. doi:10.1109/JSAC.2025.3543516.

[8] Fu, S., Wei, W., Yin, L. and Zhao, L. (2025) Joint optimization of 3-d placement and transmission power for a relay based covert communication system. *IEEE Transactions on Communications* : 1–1doi:10.1109/TCOMM.2025.3585509.

[9] Wu, D., Sun, X. and Ansari, N. (2020) An fso-based drone assisted mobile access network for emergency communications. *IEEE Transactions on Network Science and Engineering* **7**(3): 1597–1606. doi:10.1109/TNSE.2019.2942266.

[10] Wu, D., Sun, X. and Ansari, N. (2020) An fso-based drone charging system for emergency communications. *IEEE Transactions on Vehicular Technology* **69**(12): 16155–16162. doi:10.1109/TVT.2020.3043969.

[11] Zhang, S. and Ansari, N. (2021) Latency aware 3d placement and user association in drone-assisted heterogeneous networks with fso-based backhaul. *IEEE Transactions on Vehicular Technology* **70**(11): 11991–12000. doi:10.1109/TVT.2021.3112483.

[12] Lee, J.H., Park, K.H., Ko, Y.C. and Alouini, M.S. (2022) Spectral-efficient network design for high-altitude platform station networks with mixed rf/fso system. *IEEE Transactions on Wireless Communications* **21**(9): 7072–7087. doi:10.1109/TWC.2022.3154401.

[13] Parvaresh, N., Kulhandjian, M., Kulhandjian, H., D'Amours, C. and Kantarci, B. (2022) A tutorial on ai-powered 3d deployment of drone base stations: State of the art, applications and challenges. *Vehicular Communications* **36**: 100474. doi:https://doi.org/10.1016/j.vehcom.2022.100474, URL https://www.sciencedirect.com/science/article/pii/S2214209622000213.

[14] Yu, L., Sun, X., Shao, S., Chen, Y. and Albelaihi, R. (2023) Backhaul-aware drone base station placement and resource management for fso-based drone-assisted mobile networks. *IEEE Transactions on Network Science and Engineering* **10**(3): 1659–1668. doi:10.1109/TNSE.2022.3233004.

[15] Guan, Y., Zou, S., Peng, H., Ni, W., Sun, Y. and Gao, H. (2024) Cooperative uav trajectory design for disaster area emergency communications: A multiagent ppo method. *IEEE Internet of Things Journal* **11**(5): 8848–8859. doi:10.1109/JIOT.2023.3320796.

[16] Liu, S., Dahrouj, H. and Alouini, M.S. (2024) Joint user association and beamforming in integrated satellite-haps-ground networks. *IEEE Transactions on Vehicular Technology* **73**(4): 5162–5178. doi:10.1109/TVT.2023.3329168.

[17] Nguyen, T.V., Le, H.D., Mai, V., R., S. and Pham, A.T. (2024) Deep reinforcement learning for uav placement over mixed fso/rf-based non-terrestrial networks. In *2024 IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS)*: 1–5. doi:10.1109/APWCS61586.2024.10679285.

[18] Pham, P.D., Nguyen, C.K.P., Le, H.D., Pham, H.T.T., Nguyen, T.V. and Dang, N.T. (2024) Optical intelligent reflecting surface-assisted multiple users over turbulence channels. In *2024 RIVF International Conference on Computing and Communication Technologies (RIVF)*: 45–50. doi:10.1109/RIVF64335.2024.11009028.

[19] Le, H.D. and Pham, A.T. (2021) Level crossing rate and average fade duration of satellite-to-uav fso channels. *IEEE Photonics Journal* **13**(1): 1–14. doi:10.1109/JPHOT.2021.3057198.

[20] Nguyen, T.V., Le, H.D., Dang, N.T. and Pham, A.T. (2021) Average transmission rate and outage performance of relay-assisted satellite hybrid fso/rf systems. In *2021 International Conference on Advanced Technologies for Communications (ATC)*: 1–6. doi:10.1109/ATC52653.2021.9598287.

[21] Do, T.H., Nguyen, H.Q., Le, H.D., Pham, H.T.T., Nguyen, T.V. and Dang, N.T. (2024) Mac protocols for uav-aided relaying optical ground-to-hap networks. In *2024 International Conference on Advanced Technologies for Communications (ATC)*: 992–997. doi:10.1109/ATC63255.2024.10908181.

[22] Le, H.D., Nguyen, T.V. and Pham, A.T. (2021) Cloud attenuation statistical model for satellite-based fso communications. *IEEE Antennas and Wireless Propagation Letters* **20**(5): 643–647. doi:10.1109/LAWP.2021.3058641.

[23] Nguyen, T., Le, H., Pham, H. and Dang, N. (2023) Availability of free-space laser communication link with the presence of clouds in tropical regions. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems* **10**(3). doi:10.4108/eetinis.v10i3.3327.

[24] Yahia, O.B., Erdogan, E., Kurt, G.K., Altunbas, I. and Yanikomeroglu, H. (2022) Haps selection for hybrid rf/fso satellite networks. *IEEE Transactions on Aerospace and Electronic Systems* **58**(4): 2855–2867. doi:10.1109/TAES.2022.3142116.

[25] Nguyen, T.V., Pham, T.V., Dang, N.T. and Pham, A.T. (2020) Uav-based fso systems using sc-qam signaling over fading channels with misalignment. In *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*: 1–5. doi:10.1109/VTC2020-Fall49728.2020.9348866.

[26] Le, H.D. and Pham, A.T. (2022) On the design of fso-based satellite systems using incremental redundancy hybrid arq protocols with rate adaptation. *IEEE Transactions on Vehicular Technology* **71**(1): 463–477. doi:10.1109/TVT.2021.3127193.

[27] Samy, R., Yang, H.C., Rakia, T. and Alouini, M.S. (2022) Ergodic capacity analysis of satellite communication systems with sag-fso/sh-fso/rf transmission. *IEEE Photonics Journal* **14**(5): 1–9. doi:10.1109/JPHOT.2022.3201046.

[28] Wang, H., Zhang, Z., Zhu, B., Dang, J., Wu, L. and Zhang, Y. (2022) Approaches to array-type optical irss: Schemes and comparative analysis. *Journal of Lightwave Technology* **40**(12): 3576–3591. doi:10.1109/JLT.2022.3152812.

[29] Camp, T., Boleng, J. and Davies, V. (2002) A survey of mobility models for ad hoc network research.

Wireless Communications and Mobile Computing **2**. doi:10.1002/wcm.72.

[30] Chu, T.M.C., Zepernick, H.J. and Duong, T.Q. (2022) Noma-based full-duplex uav network with k-means clustering for disaster scenarios. In *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*: 1–7. doi:10.1109/VTC2022-Fall57202.2022.10012723.

[31] Khayat, G., Mavromoustakis, C.X., Pitsillides, A., Batalla, J.M. and Markakis, E.K. (2023) Multiple redundant k-means clustered scheme based on weighted cluster head selection for damaged s-uav. In *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*: 6177–6182. doi:10.1109/GLOBECOM54140.2023.10437501.

[32] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A. *et al.* (2015) Human-level control through deep reinforcement learning. *Nature* **518**(7540): 529–533. doi:10.1038/nature14236.