

Quantum DRL-Based UAV Placement with Optical Sensing-Enabled Hybrid FSO/RF Backhaul Design

Hung V. Le*, Hoang D. Le*, Duy-Tuan Dao[†], and Anh T. Pham*

* Computer Communications Laboratory, The University of Aizu, Aizuwakamatsu 965-8580, Japan

[†] The University of Danang-University of Science and Technology, Da Nang 50000, Vietnam

Abstract—This study introduces an architecture integrating optical sensing-enabled hybrid free-space optics (FSO)/radio-frequency (RF) backhaul networks with unmanned aerial vehicle (UAV)-mounted base stations to provide high-speed connectivity for ground mobile users (MUs). To maximize the number of served high-data-rate MUs, a backhaul link-switching scheme is adopted to switch between FSO and RF links under varying environmental conditions, while a quantum deep reinforcement learning (QDRL) model is proposed for optimal UAV placement. Simulation results demonstrate the effectiveness of the link-switching scheme in maintaining stable backhaul connectivity under dynamic environmental conditions. Notably, the proposed QDRL outperforms classical DRL benchmarks in terms of learning stability and parameter efficiency, highlighting its potential to provide stable control policies and scalable solutions for dynamic and large-scale networks.

Index Terms—Optical sensing, free-space optics (FSO), quantum deep reinforcement learning (QDRL).

I. INTRODUCTION

In beyond-5G networks, UAV-assisted aerial networks equipped with free-space optics (FSO)-based backhaul links offer extended coverage, deployment flexibility, and high-capacity data transmission [1], [2]. However, FSO links are highly susceptible to environmental conditions such as fog. To address this challenge, an optical sensing-assisted FSO/RF backhaul link switching scheme is a promising solution, in which optical sensing rapidly monitors environmental conditions [3] to enable adaptive switching between FSO and RF links under different weather conditions. Nevertheless, varying conditions in both backhaul and access links remain a major challenge for network optimization. Deep reinforcement learning (DRL) is well suited for dynamic networks because of its fast inference capability. Nevertheless, DRL models require a large number of trainable parameters because they employ deep neural networks (DNNs) [4]–[6]. Especially in the context of network optimization, the number of parameters increases as the network scales up. To address this challenge, quantum deep reinforcement learning (QDRL) has emerged as a parameter-efficient solution that exploits quantum superposition and entanglement to achieve expressive policy representations with fewer trainable parameters [5], [6].

Most existing studies have employed DRL for UAV placement optimization [7]–[9]. Notably, Prabhashana et al. [10] first proposed a quantum-enhanced actor-critic model for UAV placement in a 6G-enabled SAGIN. In addition, Nafees et al. [11] first introduced Q-learning for UAV trajectory optimization in mixed FSO/RF networks using optical sensing to sense

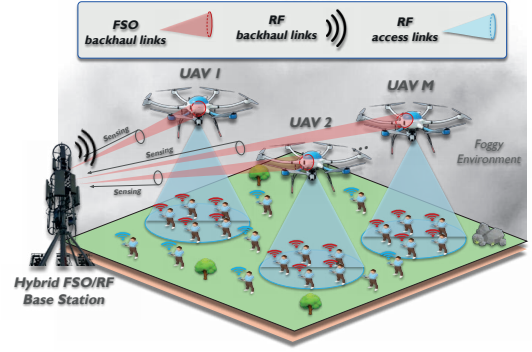


Fig. 1: The considered optical sensing-enabled hybrid FSO/RF backhaul network

fog conditions. To the best of our knowledge, QDRL-based optimal multi-UAV placement with an optical sensing-enabled hybrid FSO/RF backhaul design has not yet been explored in the existing literature. Our study proposes a QDRL model for optimal UAV placement with an optical sensing-enabled hybrid FSO/RF backhaul design.

II. SYSTEM MODEL

The considered system depicted in Fig. 1 consists of an optical sensing-assisted ground base station that provides hybrid FSO/RF backhaul links to UAVs serving mobile users (MUs) within a target area. An optical sensing system is installed at the base station to transmit FSO signals to the UAVs while simultaneously sensing fog conditions through the reflected optical signals [11]. For backhaul communications assisted by optical fog sensing, the base station dynamically switches between FSO and RF links, using FSO under light fog conditions and RF when heavy fog degrades optical transmission. In the access network, UAV-mounted base stations provide high-speed RF connectivity to ground MUs.

The UAV operating time is partitioned into T equal intervals, each of duration τ , selected such that the network can be assumed quasi-static within each interval. We consider a system where M UAVs are deployed to serve MUs within a target region containing N MUs. The indices of the UAVs, MUs, and time slots are denoted by $m \in \mathcal{M} = \{1, 2, \dots, M\}$, $n \in \mathcal{N} = \{1, 2, \dots, N\}$, and $t \in \mathcal{T} = \{1, 2, \dots, T\}$, respectively. The m -th UAV operates at altitude $h_{m,t}^{\text{UAV}}$ and moves with a fixed velocity v_m . The optical sensing system

of the base station located at $(0, 0)$ is installed at an altitude of h^{BS} . The coordinate of the m -th UAV is $(x_{m,t}^{\text{UAV}}, y_{m,t}^{\text{UAV}})$. The location of the n -th MU is $(x_{n,t}^{\text{MU}}, y_{n,t}^{\text{MU}})$. The mobility of the MUs is modeled using the Gauss–Markov mobility model [12].

A. Backhaul Channel

For the FSO backhaul channel of the m -th UAV, we consider two major channel impairments including geometric loss $h_{m,t}^{\text{geo}}$ and atmospheric loss $h_{m,t}^{\text{at}}$. The composite channel gain is expressed as $h_{m,t} = h_{m,t}^{\text{geo}} h_{m,t}^{\text{at}}$, where $h_{m,t}^{\text{geo}}$ and $h_{m,t}^{\text{at}}$ are expressed by

$$h_{m,t}^{\text{geo}} = \text{erf} \left(\frac{\sqrt{\pi}\gamma}{2\sqrt{2}\phi L_{m,t}^{\text{UAV}}} \right)^2; h_{m,t}^{\text{at}} = \exp(-\vartheta L_{m,t}^{\text{UAV}}) \quad (1)$$

where γ denotes the receiver aperture diameter. ϕ is divergence angle. $L_{m,t}^{\text{UAV}} = \sqrt{x_{m,t}^{\text{UAV}2} + y_{m,t}^{\text{UAV}2} + (h_{m,t}^{\text{UAV}} - h^{\text{BS}})^2}$ is the distance between FSO transmitter and the m -th UAV. $\vartheta = \frac{3.91}{V_{m,t}} \left(\frac{\lambda_m}{550 \times 10^{-9}} \right)^{-\epsilon}$ is the attenuation coefficient, where $V_{m,t}$ is the visibility, and ϵ represents the weather condition-dependent size distribution of scattering particles [11]. Given $h_{m,t}$, the FSO backhaul capacity of the m -th UAV is determined by

$$C_{m,t}^{\text{FSO}} = \frac{1}{2} B_m^{\text{FSO}} \log_2 (1 + \gamma_{m,t}^{\text{FSO}}); \gamma_{m,t}^{\text{FSO}} = \frac{e P_{m,t}^{\text{FSO}2} h_{m,t}^2 \rho^2}{2\pi \sigma_{\text{UAV,FSO}}^2} \quad (2)$$

where B_m^{FSO} is the FSO bandwidth allocated to the m -th UAV, e is the base of natural logarithm, $P_{m,t}^{\text{FSO}}$ is the optical transmit power for the m -th UAV. ρ is the photodetector's responsivity. $\sigma_{\text{UAV,FSO}}^2$ is the noise variance.

As for the RF backhaul channel for the m -th UAV, the path loss of m -th UAV is expressed as $\mathcal{L}_{m,t} = \left(\frac{4\pi f_m L_{m,t}^{\text{UAV}}}{c} \right)^\eta$. Given the path loss $\mathcal{L}_{m,t}$, the RF backhaul capacity of the m -th UAV is determined by

$$C_{m,t}^{\text{RF}} = B_m^{\text{RF}} \log_2 \left(1 + \frac{P_{m,t}^{\text{RF}} G_t^m G_r^m / \mathcal{L}_{m,t}}{\sigma_{\text{UAV,RF}}^2} \right) \quad (3)$$

where B_m^{RF} is the RF bandwidth allocated to the m -th UAV, $P_{m,t}^{\text{RF}}$ is the RF transmit power of the m -th UAV. G_t and G_r are the transmit and receive antenna gains of the m -th UAV, respectively. $\sigma_{\text{UAV,RF}}^2$ is the noise variance.

B. Optical Sensing-Enabled Hybrid FSO/RF Link Switching

An optical sensing-assisted hybrid FSO/RF link switching scheme is adopted, in which an optical sensing system is employed to monitor fog conditions and facilitate adaptive switching between FSO and RF backhaul links. Specifically, (i) the base station transmits optical laser beams toward the UAVs; upon encountering fog droplets, a portion of the optical energy is backscattered. (ii) The backscattered optical energy is collected by a sensor located at the base station. (iii) The sensor converts the received optical energy into an electrical signal. (iv) The visibility $\hat{V}_{m,t}$ is then estimated from the electrical signal [11]. However, the estimated visibility is

assumed to be imperfect and is expressed as $\hat{V}_{m,t} = V_{m,t} + n_{\text{es}}$, where n_{es} is the estimation noise [11] following a Gaussian distribution, i.e., $n_{\text{es}} \sim \mathcal{N}(0, \sigma_{\text{es}}^2)$. Therefore, the estimated composite channel gain is expressed as $\hat{h}_{m,t} = h_{m,t}^{\text{geo}} \hat{h}_{m,t}^{\text{at}}$, where $\hat{h}_{m,t}^{\text{at}}$ is calculated from the estimated visibility $\hat{V}_{m,t}$ according to (1). The estimated SNR, $\hat{\gamma}_{m,t}^{\text{FSO}}$, is computed from the estimated channel gain $\hat{h}_{m,t}$. Based on the estimated SNR $\hat{\gamma}_{m,t}^{\text{FSO}}$ obtained through optical sensing, the link-switching scheme is defined as follows:

$$C_{m,t} = \begin{cases} C_{m,t}^{\text{FSO}}, & \text{if } \gamma_{\text{thres}} \leq \hat{\gamma}_{m,t}^{\text{FSO}}, \\ C_{m,t}^{\text{RF}}, & \text{if } \hat{\gamma}_{m,t}^{\text{FSO}} < \gamma_{\text{thres}}, \end{cases} \quad (4)$$

where γ_{thres} is a threshold used to determine link switching, $C_{m,t}$ denotes the selected backhaul capacity after link switching.

As for fog modeling, the fog condition is characterized by the visibility $V_{m,t}$, which is assumed to be uniform across the entire target area (i.e., $V_{m,t} = V_{\text{area}}$). The visibility varies over T time slots and is randomly increased or decreased to emulate dynamic fog conditions, thereby modeling the temporal variation of fog intensity.

C. Access Channel

The achievable data rate of the n -th MU served by the m -th UAV is determined by

$$R_{n,m,t}^{\text{RF}} = B_{\text{MU}}^{\text{RF}} \log_2 \left(1 + \frac{P_{m,n,t}^{\text{UAV}} g_{m,n,t}^{\text{UAV}}}{\sigma_{\text{MU,RF}}^2} \right), \quad (5)$$

where $B_{\text{MU}}^{\text{RF}}$ is the RF bandwidth allocated per MU, $P_{m,n,t}^{\text{UAV}}$ is the transmission power from the m -th UAV to the n -th MU, $\sigma_{\text{MU,RF}}^2$ is the noise variance, and channel gain $g_{m,n,t}^{\text{UAV}}$ is reported in [12, Sec. II-C].

Additionally, a user association algorithm with a power allocation scheme, as reported in [12, Sec. II-D], is adopted to adaptively allocate transmit power $P_{m,n,t}^{\text{UAV}}$ while ensuring a fixed target data rate $R_{\text{target}}^{\text{RF}}$ for MUs. The set of MUs within the coverage of the m -th UAV $\mathcal{N}_{m,t}$ is computed using the equation in [12, (11)]. The set of served MUs $\mathcal{S}_{m,t}$ is computed using the algorithm reported in [12, Algo. 1].

III. PROPOSED QDRL-BASED FRAMEWORK

A. Problem Formulation

Our objective is to find the optimal coordinate of UAVs $\mathcal{K} = \{(x_{m,t}^{\text{UAV}}, y_{m,t}^{\text{UAV}})\}_{m \in \mathcal{M}, t \in \mathcal{T}}$ to maximize the total number of served MUs. Our optimization problem $\mathcal{P}$ is formulated as

$$\mathcal{P} : \max_{\mathcal{K}} Q = \sum_{t \in \mathcal{T}} \sum_{m \in \mathcal{M}} |\mathcal{S}_{m,t}|, \quad (6a)$$

$$\text{s.t. } |\mathcal{S}_{m,t}| R_{\text{target}}^{\text{RF}} \leq C_{m,t}, \forall m, t, \quad (6b)$$

$$\sum_{n \in \mathcal{S}_{m,t}} P_{m,n,t}^{\text{UAV}} \leq P_{m,\text{max}}^{\text{UAV}}, \forall m, t, \quad (6c)$$

$$x_{\min} \leq x_{m,t}^{\text{UAV}} \leq x_{\max}, \forall m, t, \quad (6d)$$

$$y_{\min} \leq y_{m,t}^{\text{UAV}} \leq y_{\max}, \forall m, t, \quad (6e)$$

Algorithm 1 Amplitude Encoding

Require: Classical State $o_{m,t}$

Ensure: Quantum State $|\psi_{m,t}\rangle$

- 1: Initialize q qubits
- 2: **if** $|o_{m,t}| < 2^q$ **then**
- 3: Create a zero vector $v_0 = [0, 0, \dots, 0] \in \mathbb{R}^{2^q - |o_{m,t}|}$
- 4: Zero padding $o_{m,t} \leftarrow [o_{m,t}, v_0]$
- 5: **end if**
- 6: Normalize the classical state: $\mathbf{v} = \frac{o_{m,t}}{\|o_{m,t}\|_2}$
- 7: Encode $o_{m,t}$ into $|\psi_{m,t}\rangle = \sum_{i=0}^{2^q-1} v_i |i\rangle$

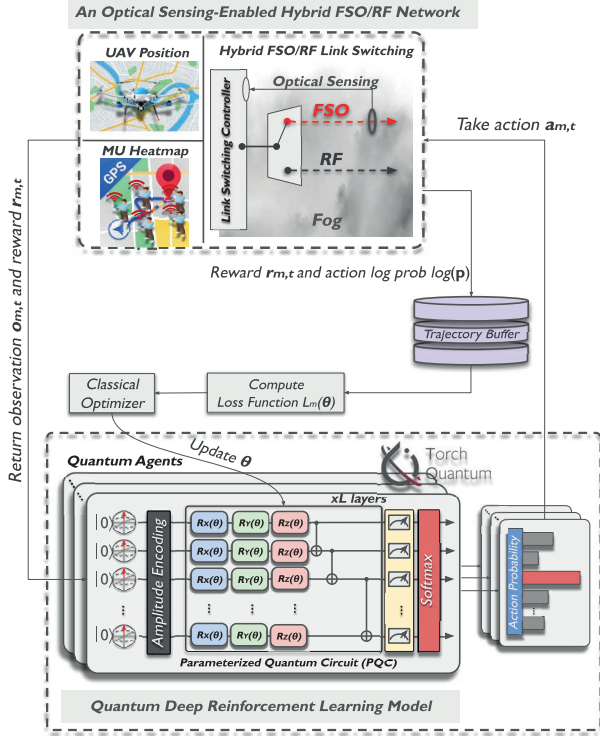


Fig. 2: Proposed QDRL-based framework for optical sensing-enabled hybrid FSO/RF backhaul networks

where (6b) represents the backhaul-capacity constraint of the UAVs, and (6c) represents the maximum transmit-power constraint of the UAVs. (6d) and (6e) denote the constraints on the UAVs' coordinates.

B. MDP Formulation

1) *Observation Space*: Let $o_{m,t}$ denote the observation of the m -th UAV in the time slot t , including the coordinates of all UAVs $\{(x_{m,t}^{\text{UAV}}, y_{m,t}^{\text{UAV}})\}_{m \in \mathcal{M}}$, the heatmap of unserved MUs and MUs within the coverage of the m -th UAV $H_{m,t} = H(\{(x_{n,t}^{\text{MU}}, y_{n,t}^{\text{MU}})\}_{n \in S_{m,t} \cup \mathcal{M}}) + H(\{(x_{n,t}^{\text{MU}}, y_{n,t}^{\text{MU}})\}_{n \in \mathcal{N}_{m,t}})$, the current backhaul capacity $C_{m,t}$. The $o_{m,t}$ is expressed as

$$o_{m,t} = [\{(x_{m,t}^{\text{UAV}}, y_{m,t}^{\text{UAV}})\}_{m \in \mathcal{M}}, H_{m,t}, C_{m,t}] \quad (7)$$

2) *Action Space*: The movement of each UAV is modeled as a set of discrete actions, including: “0” for staying station-

ary, “1” for moving up, “2” for moving down, “3” for moving left, and “4” for moving right, i.e., $a_{m,t} \in \{0, 1, 2, 3, 4\}$.

3) *Reward*: The reward of the m -th UAV is defined as the number of served MUs, i.e., $r_{m,t} = \zeta \sum_{m \in \mathcal{M}} |S_{m,t}| + (1 - \zeta)|S_{m,t}|$, where ζ is the weight ratio.

C. Proposed QDRL-based framework

To address problem $\mathcal{P}$, we propose a QDRL-based optimization framework built upon a quantum policy gradient model [13], as depicted in Fig. 2.

1) *Amplitude Encoding*: The classical state $o_{m,t}$ is encoded into the quantum state $|\psi_{m,t}\rangle$ via amplitude encoding [14], as detailed in Algorithm 1. This approach allows the representation of 2^q classical values using only q qubits. The state $o_{m,t}$ is encoded as follows: (i) choose the number of qubits q such that $|o_{m,t}| \leq 2^q$; (ii) if $|o_{m,t}| < 2^q$, perform zero-padding to obtain a vector of length 2^q ; (iii) normalize the vector to $\mathbf{v}$ such that $\|\mathbf{v}\|_2^2 = 1$; and (iv) map the normalized vector to the quantum state $|\psi_{m,t}\rangle$.

2) *Parameterized Quantum Circuit*: A parameterized quantum circuit (PQC) is structured as a network of multiple quantum gates. The quantum-based policy network for quantum agent m employs a PQC network U_θ^m , consisting of a sequence of L similar PQC blocks. The policy network U_θ^m , structured with rotation layers $U_{\text{rot}}^{(l)}$ and entangling layers $U_{\text{ent}}^{(l)}$, is expressed as

$$U_\theta^m = \prod_{l=1}^L U_{\text{ent}}^{(l)} U_{\text{rot}}^{(l)} \quad (8)$$

where $U_{\text{rot}}^{(l)} = \bigotimes_{i=1}^q \left(R_Z(\theta_{i,l}^z) R_Y(\theta_{i,l}^y) R_X(\theta_{i,l}^x) \right)$ represents a sequence of R_X , R_Y , and R_Z rotation gates applied to q qubits. $U_{\text{ent}}^{(l)} = \prod_{i=1}^{q-1} \text{CNOT}_{i,i+1}$ represents a sequence of CNOT gates applied to adjacent qubit pairs. Therefore, the full expression of U_θ^m is given as

$$U_\theta^m = \prod_{l=1}^L \left[\left(\prod_{i=1}^{q-1} \text{CNOT}_{i,i+1} \right) \times \left(\bigotimes_{i=1}^q R_Z(\theta_{i,l}^z) R_Y(\theta_{i,l}^y) R_X(\theta_{i,l}^x) \right) \right] \quad (9)$$

3) *Measurement*: After passing the quantum state $|\psi_{m,t}\rangle$ through the PQC U_θ^m , the resulting output state is expressed as $|\phi_{m,t}\rangle = U_\theta^m |\psi_{m,t}\rangle$. To extract classical information from the quantum state $|\phi_{m,t}\rangle$, quantum measurements are performed using observables. Specifically, the expectation values of the Pauli-Z observables measured on q qubits are collected to form the vector $\mathbf{z} = [\langle \phi_{m,t} | Z_i | \phi_{m,t} \rangle]_{i=1}^q$.

4) *Training Algorithm*: Our proposed QDRL is trained using the REINFORCE training algorithm [13]. The training procedure of the proposed QDRL is detailed in Algorithm 2.

IV. SIMULATION RESULTS AND DISCUSSION

A. Simulation Settings

The parameter settings for the backhaul, access links, proposed QDRL model, and benchmark models are presented as follows.

TABLE I: Number of trainable parameters of the proposed QDRL and classical DRL models

Algorithm	Number of parameters	Parameter ratio
PPO	42246×3	100%
DQN	21381×3	50.611%
Proposed QDRL	120×3	0.284%

Algorithm 2 Proposed QDRL model's training algorithm

```

1: Initialize parameters  $\theta$  for PQC  $U_\theta^m$  for quantum agent  $m$ 
2: Initialize the discount factor  $\gamma_{DF}$  and learning rate  $\alpha_{LR}$ 
3: for each episode 1 to  $E$  do
4:   Initialize a trajectory buffer  $\mathcal{D} = []$ 
5:   for step 1 to  $T$  do
6:     Encode  $o_{m,t}$  into  $|\psi_{m,t}\rangle$  using Algorithm.1
7:     Apply  $U_\theta^m$  to  $|\psi_{m,t}\rangle$ :  $|\phi_{m,t}\rangle = U_\theta^m |\psi_{m,t}\rangle$ 
8:     Measure  $\mathbf{z} = [\langle \phi_m | Z_i | \phi_m \rangle]_{i=1}^q$  using Pauli-Z
9:     Transform into probability vector  $\mathbf{p} = \text{Softmax}(\mathbf{z})$ 
10:    Take action  $a_{m,t} \sim \mathbf{p}$ 
11:     $\mathcal{D} \leftarrow \mathcal{D} \cup [r_{m,t}, \log \mathbf{p}]$ 
12:   end for
13:   Initial return  $G_0 = 0$ 
14:   for all samples stored in  $\mathcal{D}$  do
15:     Discounted return  $G_t = \sum_{k=1}^t \gamma_{DF}^k r_{m,k}$ 
16:     Loss  $\mathcal{L}_m(\theta) = -\sum_{t=1}^T G_t \log \mathbf{p}$ 
17:   end for
18:   Update  $\theta \leftarrow \theta + \alpha_{LR} \nabla_\theta \mathcal{L}_m(\theta)$  using Adam optimizer
19: end for

```

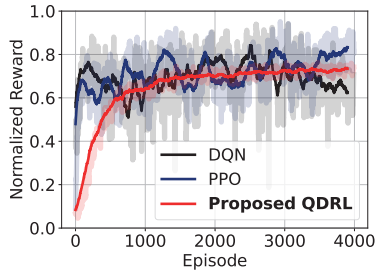


Fig. 3: Comparison of reward convergence between our proposed QDRL and other classical DRL models.

1) *Backhaul links*: For FSO links, transmit power $P_{m,t}^{\text{FSO}} = 1$ W, divergence angle $\phi = 0.001$ rad, optical wavelengths $\lambda_1 = 1550$ nm, $\lambda_2 = 1555$ nm, $\lambda_3 = 1560$ nm, noise power $\sigma_{\text{UAV,FSO}}^2 = 10^{-6}$ W, FSO bandwidth $B_m^{\text{FSO}} = 1$ GHz. As for RF links, transmit power $P_{m,t}^{\text{RF}} = 50$ W, RF bandwidth $B_m^{\text{RF}} = 1$ MHz, noise power $\sigma_{\text{UAV,RF}}^2 = 10^{-4}$ W, frequencies $f_1 = 2$ GHz, $f_2 = 2.1$ GHz, $f_3 = 2.2$ GHz. The visibility range is [0.05,10] km.

2) *Access links*: $M = 3$ UAVs starting at $(x_{m,0}^{\text{UAV}}, y_{m,0}^{\text{UAV}}) = (0, 0)$ are deployed to serve MUs in an area of $2 \text{ km} \times 2 \text{ km}$. The altitude of UAV $h_{m,t}^{\text{UAV}} = 350$, the UAV velocity is $v_m = 20$ m/s. Number of MUs $N = 350$, their locations follow a Gaussian distribution $\mathcal{N}(\mu = 1000, \sigma^2 = 300)$. The heatmap of MUs is represented as a 5×5 grid.

3) *QDRL*: The observation vector $o_{m,t}$ is 32-dimensional and is encoded using $q = 5$ qubits. Number of PQC blocks $L = 8$. The QDRL model is implemented using the open-source TorchQuantum framework. Number of episodes $E = 4000$, number of steps $T = 300$, discount factor $\gamma_{DF} = 0.99$. The classical optimizer is Adam with a learning rate of $\alpha_{LR} = 0.0001$. Each quantum agent controls a UAV, and all agents are trained using decentralized training and execution.

4) *Benchmarks*: To evaluate the effectiveness of the proposed QDRL framework, 2 classical DRL models, namely PPO and DQN, are used as benchmarks. DQN is implemented using a DNN-based value network with layer sizes $\{32, 128, 128, 5\}$. Similarly, PPO employs a DNN-based actor network with layer sizes $\{32, 128, 128, 5\}$ and a DNN-based critic network with layer sizes $\{32, 128, 128, 1\}$. Both DQN and PPO are trained using decentralized training and execution.

B. Simulation Results

Firstly, we compare the performance of the proposed QDRL with classical DRL models. Figure 3 shows that the proposed QDRL achieves reward convergence performance comparable to those of DQN and PPO. Notably, the proposed QDRL demonstrates a smoother and more stable convergence trajectory than DQN and PPO, indicating superior learning stability and a more reliable policy. Consequently, it has the potential to enable more consistent and reliable UAV control decisions in dynamic networks. In addition, Table I compares the number of trainable parameters required by the proposed QDRL, DQN, and PPO models. While DQN and PPO require 21381×3 and 42246×3 parameters, respectively, the proposed QDRL requires only 120×3 parameters, demonstrating a substantial reduction in model complexity.

Secondly, we analyze the performance of optical sensing-enabled hybrid FSO/RF link switching. Figure 4 illustrates the trajectories of the three UAVs and the corresponding backhaul link selection under different atmospheric conditions, namely clear weather, haze, and dense fog. As visibility decreases from clear weather to dense fog, FSO links become less reliable, prompting the UAVs to switch to RF backhaul links. Specifically, under clear weather conditions, as depicted in Fig. 4a, FSO backhaul links are employed for all three UAVs. Under haze conditions, as shown in Fig. 4b, only UAV 1 is served by an FSO backhaul link, while the other two UAVs switch to RF backhaul links due to their greater distances from the base station. Under dense fog conditions, as shown in Fig. 4c, all UAVs are connected via RF backhaul links because the FSO links are severely degraded. Figure 5 depicts the number of served MUs over time under decreasing visibility. The proposed hybrid FSO/RF scheme maintains a stable number of served MUs by switching from FSO to RF links when visibility decreases. In contrast, without link switching, the UAVs continue using FSO backhaul links, causing the number of served MUs to drop to zero.

V. CONCLUSION

This work introduced an architecture that integrates an optical sensing-enabled hybrid FSO/RF backhaul network

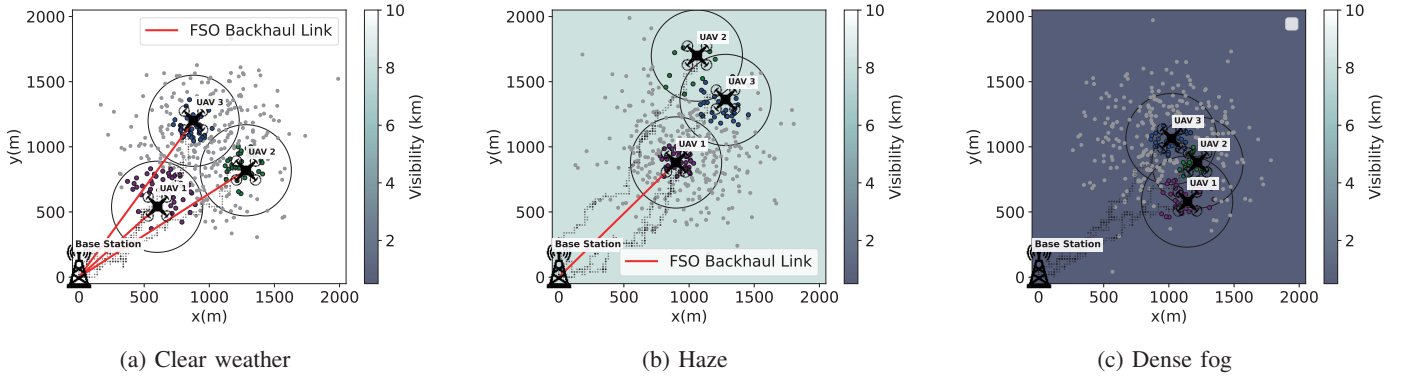


Fig. 4: Illustration of UAV trajectories and hybrid FSO/RF backhaul link switching under different fog conditions.

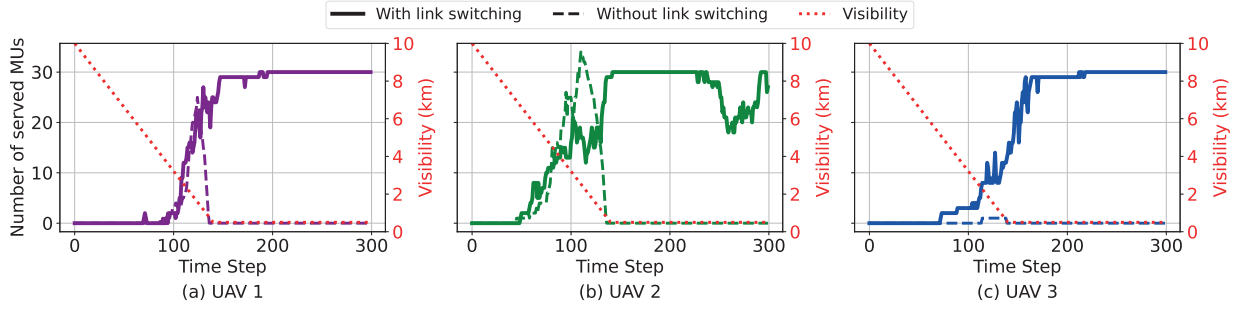


Fig. 5: Number of served MUs by UAVs with and without FSO/RF link switching.

with UAV-mounted base stations for RF access networks. An optical sensing-enabled link-switching scheme was adopted for the hybrid backhaul, while a QDRL model was employed for multi-UAV placement optimization. The simulation results showed that the proposed QDRL achieved superior learning stability with significantly fewer trainable parameters than conventional DRL models, highlighting the potential of quantum-enhanced DRL as a stable and scalable approach for optimizing dynamic and large-scale networks. In addition, the effectiveness of the adaptive backhaul link-switching scheme in maintaining backhaul capacity under varying environmental conditions was demonstrated.

ACKNOWLEDGMENT

This work was supported by the Telecommunications Advancement Foundation (TAF).

REFERENCES

- [1] Y. Guan, S. Zou, H. Peng, W. Ni, Y. Sun, and H. Gao, "Cooperative uav trajectory design for disaster area emergency communications: A multiagent ppo method," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 8848–8859, 2023.
- [2] F. H. Panahi and F. H. Panahi, "Cellular coverage extension using an intelligent fso-based uav: An energy and spectral efficient approach," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 1, pp. 556–565, 2025.
- [3] Y. Wen, F. Yang, J. Song, and Z. Han, "Optical integrated sensing and communication: Architectures, potentials and challenges," *IEEE Internet Things Mag.*, vol. 7, no. 4, pp. 68–74, 2024.
- [4] G. S. Kim, S. Lee, I.-S. Cho, S. Park, and J. Kim, "Quantum Reinforcement Learning for Lightweight LEO Satellite Routing," *IEEE Internet Things J.*, 2025.
- [5] A. A. Puspitasari and B. M. Lee, "Quantum reinforcement learning for uav-enhanced next-generation wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 28, pp. 3089–3124, 2025.
- [6] N. M. E. P. Astiti and B. M. Lee, "A survey of quantum deep reinforcement learning for resource allocation in future wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 28, pp. 109–148, 2025.
- [7] H. Zhang, Z. Qi, J. Li, A. Aronsson, J. Bosch, and H. H. Olsson, "5g network on wings: A deep reinforcement learning approach to the uav-based integrated access and backhaul," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 2, pp. 1109–1126, 2024.
- [8] Y. Bai, B. Xie, Y. Liu, Z. Chang, and R. Jäntti, "Dynamic uav deployment in multi-uav wireless networks: A multi-modal feature-based deep reinforcement learning approach," *IEEE Internet Things J.*, 2025.
- [9] X. Liu, J. Wu, C. Zhao, and Z. Liu, "Integrated sensing and communications for uav assisted internet of things based on deep reinforcement learning," *IEEE Trans. Veh. Technol.*, 2025.
- [10] S. C. Prabhashana, D. V. Huynh, O. A. Dobre, H. Shin, and T. Q. Duong, "Quantum drl for green uav positioning in 6g-enabled sagin with cooperative nano-satellite constellations," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2025, pp. 6484–6489.
- [11] M. Nafees, M. Baniyadi, J. R. Hopgood, M. Safari, and J. S. Thompson, "Integrated sensing and communication for uav trajectory optimization in mixed fso-rf networks in dynamic weather conditions," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, IEEE, 2025, pp. 1–6.
- [12] H. V. Le, H. D. Le, D.-T. Dao, and A. T. Pham, "Multi-agent deep reinforcement learning for uav placement in optical irs-aided aerial networks," in *Proc. 30th Asia-Pacific Conf. Commun. (APCC)*, 2025, pp. 1–6.
- [13] S. Jerbi, C. Gyurik, S. Marshall, H. Briegel, and V. Dunjko, "Parametrized quantum policies for reinforcement learning," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 28 362–28 375, 2021.
- [14] A. Paul, K. Singh, A. Kaushik, C.-P. Li, O. A. Dobre, M. D. Renzo, and T. Q. Duong, "Quantum-Enhanced DRL Optimization for DoA Estimation and Task Offloading in ISAC Systems," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 1, pp. 364–381, 2024.