D3QN-Based Trajectory and Handover Management for UAVs Co-existing with Terrestrial Users

Yuhang Deng, Irshad A. Meer, Shuai Zhang, Mustafa Ozger, and Cicek Cavdar
School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden
E-mail: {yuhangd, iameer, shuai2, ozger, cavdar}@kth.se

***Abstract*—The ubiquitous cellular network is a strong candidate for providing UAVs' wireless connectivity. Due to the maneuverability advantage and higher altitude, UAVs could have line-of-sight (LoS) connectivity with more base station (BS) candidates than terrestrial users. However, the LoS connectivity could also enhance the propagation of up-link interference caused by UAVs over co-existing terrestrial users. In addition, UAVs would perform more handovers than terrestrial users when moving due to the extensive overlap in the coverage areas of many BS candidates. The solution is to bypass the overlapping coverage areas by designing the UAVs' trajectory and to reduce interference by optimizing radio resource allocation through handover management. This paper studies the joint optimization of a UAV's trajectory design and handover management to minimize the weighted sum of three key performance indicators (KPIs): delay, up-link interference, and handover numbers. A dueling double deep Q-network (D3QN) based reinforcement learning algorithm is proposed to solve the optimization problem. Results show that the proposed approach can reduce the handover numbers by 90% and the interference by 18% at the cost of a small increment in transmission delay when compared with the benchmark scheme, which controls the UAV to move along the shortest path and perform handover based on received signal strength. Finally, we verify the advantage of introducing trajectory design, which can reduce the interference by 29% and eliminate the handover numbers by 33% when compared to the D3QN-based policy without trajectory design.**

***Index Terms*—cellular-connected UAVs, trajectory design, handover management, radio resource allocation, reinforcement learning, machine learning.**

I. Introduction

Unmanned aerial vehicles (UAVs) have the characteristics of low cost and ease of deployment, resulting in various applications [1], [2]. Leveraging ubiquitous cellular networks to provide wireless connectivity for UAVs offers a compelling cost-effective solution, eliminating the need for building communication infrastructure dedicated to UAVs [3]–[6]. However, since current cellular systems are designed for terrestrial user equipment (UE), there are some unresolved issues related to the coexistence of UAVs and the terrestrial UE [7]–[9].

Compared to terrestrial UE, UAVs flying at higher altitudes benefit from an expanded field of view, enabling them to potentially establish line-of-sight (LoS) connectivity with a greater number of base stations (BSs) [9]. Although the LoS channels can provide UAVs with stronger wireless connectivity to the serving BSs, they also increase the up-link interference from UAVs to adjacent BSs serving terrestrial UE. Therefore, the up-link interference caused by UAVs is one of the issues that remain to be solved for the coexistence of UAVs and the terrestrial UE. Furthermore, authors in [10] substantiate that UAVs tend to experience more handovers compared to terrestrial UE when in motion, with a significant number of these handovers being unnecessary. Handover decisions are conventionally performed by comparing received signal strengths (RSSs) from a set of BS candidates. Thus, handovers often occur at the edge of a cell's coverage area, where the RSSs from multiple BSs may be comparable. A small fluctuation in the RSSs from different BSs can lead to a change in the serving BS. Due to the LoS conditions between UAVs and many BSs, there can be a large overlap in the coverage areas of different BSs, causing drastic fluctuations in RSSs for UAVs while moving. Thus, performing handovers based on RSSs can cause UAVs to perform many redundant handovers.

Literature on the seamless integration of UAVs into existing cellular networks primarily focuses on two perspectives: trajectory design and handover management. However, existing literature typically optimizes these two perspectives separately. Trajectory design is a widely adopted approach since it could leverage the agility and maneuverability of UAVs to address coexistence challenges with terrestrial UE [11]–[14]. The authors in [11] improve the end-to-end throughput for a UAV-relayed wireless network by designing the trajectory of the UAV to search around blocking buildings for an optimal relay position. In [12], the authors optimize users' weighted sum rate by designing the trajectory of a UAV relay with given optional landing spots. In [13], an approach utilizing deep reinforcement learning (DRL) is proposed to design trajectories for multiple UAVs, aiming to optimize their coverage performance. An interference management approach based on DRL is proposed in [14]. The authors investigate the trade-off between the transmission delay of multiple UAVs and their up-link interference through trajectory design. The mentioned research primarily focuses on optimizing the trajectory of UAVs, while overlooking the importance of their handover management. However, handover management for UAVs has gained considerable attention due to its significance in avoiding the ping-pong effect and improving the efficiency

This work was supported in part by the CELTIC-NEXT Project, 6G for Connected Sky (6G-SKY), with funding received from Vinnova, Swedish Innovation Agency.

of wireless resource utilization for UAVs [15], [16]. In [15], the authors propose a deep Q-network (DQN) based algorithm to manage the BSs association for a UAV moving along a given path, thereby significantly reducing the redundant handovers at the cost of a slight loss in signal strength. In [16], the authors jointly optimize the delay, up-link interference, and handover numbers for a UAV with a proposed DQN algorithm. They manage the BSs association and radio resource blocks (RRBs) allocation for a UAV moving along a given path.

Considering the significance of individually optimizing trajectory design and handover management highlighted in the aforementioned works, jointly optimizing them will yield greater performance improvements. On the one hand, implementing an optimized handover management policy helps to minimize unnecessary handovers by strategically designing the BSs association and RRBs allocation for a UAV, thus reducing transmission delay and up-link interference. On the other hand, trajectory design can mitigate the up-link interference produced by the UAV and improve its throughput by making it move along the radiation direction of the serving BS and bypass the overlapping coverage areas of adjacent BSs. However, joint optimization of trajectory design and handover management to minimize three key performance indicators (KPIs), i.e., delay, up-link interference, and handover numbers, is still an open research question. The difficulty of this research question lies in its scale, which grows exponentially with the expansion of the UAV's moving range and the number of available BSs. DRL is a powerful tool for solving complex problems [17]–[19]. In this paper, we design a joint optimization approach based on a DRL algorithm, the dueling double deep Q-network (D3QN) algorithm, to solve the joint optimization problem. Our contributions are described as follows:

- We formulate a joint optimization problem to design trajectory and manage the BSs association as well as the RRBs allocation for a UAV while minimizing the three aforementioned KPIs.
- We transform the joint optimization problem into a DRL problem and propose a solution based on the D3QN algorithm, which outperforms the shortest path scheme.
- Compared with the fixed-path approach proposed by [16], our proposed solution offers the advantage of adaptively adjusting the UAV's trajectory, resulting in reduced up-link interference and fewer handovers for the UAV while ensuring a comparable transmission delay.

The remainder of the paper is organized as follows: Section II presents the basic setting of our system model. Section III formulates the optimization problem. Section IV transfers the problem into the objective of our proposed D3QN solution and gives a detailed introduction to its framework. Section V demonstrates our simulation results and the performance of our solution. Finally, we conclude the paper in Section VI.

II. System Model

In Figure 1, we examine the up-link communication scenario involving a cellular-connected UAV coexisting with terrestrial UE. In this setup, we utilize single carrier frequency division multiple access for both the UAV and terrestrial UE. The blue line indicates the direct air-to-ground (DA2G) link between the UAV and the serving BS. The red lines show the interfering links from the UAV to adjacent BSs. The green lines represent the communication links between terrestrial UE and serving BSs. The terrestrial UE would experience interference from the UAV if they utilize the same RRBs.

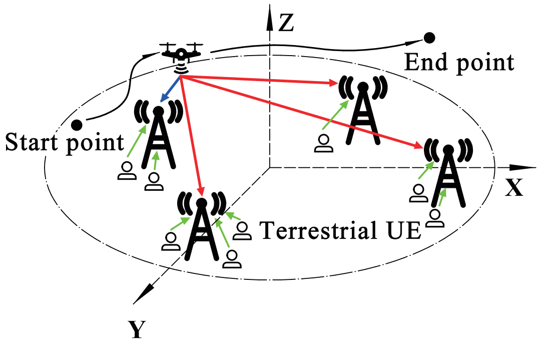

Fig. 1. The cellular network with a cellular-connected UAV.

The entire area is served by L co-channel BSs. The number of available RRBs in the system is N_b. The UAV is required to fly from a start point to an end point within the service area over T timesteps. The length of a timestep is denoted by $|t_s|$. At each timestep, the UAV must be served by one of the BSs and allocated at least one RRB from the BS. Let N_l^t denote the number of terrestrial UE served by BS l at timestep t, which is a time-correlated random variable with normal distribution. We assume that each terrestrial UE is allocated exactly one RRB and the terrestrial UE is given priority over the UAV. Thus only after satisfying the terrestrial UE's requirement could the remaining RRBs be allocated to the UAV. Finally, a machine learning (ML) agent is located in the service area as a central controller, having full access to the channel state information (CSI) about the DA2G channels between the UAV and L BSs.

A. DA2G channel model

In this paper, we adopt the probabilistic path loss model proposed in [8] for the DA2G channel. According to [20]–[22], the path loss between a UAV and a terrestrial BS depends on their distance and LoS condition. The probability of existing a LoS channel between BS l and a UAV at altitude $h \in (22.5\text{m}, 100\text{m}]$ is calculated as:

$$\mathcal{P}_{LOS} = \begin{cases} 1, & d_{\text{2D}}^l \leq d_1 \\ \frac{d_1}{d_{\text{2D}}^l} + e^{(d_1/p_1)} \cdot \left(1 - \frac{d_1}{d_{\text{2D}}^l}\right), & d_{\text{2D}}^l > d_1 \end{cases}, \tag{1}$$

where $d_1 = \max\left(\left(460\log_{10}(h) - 700\right), 800\right)$, $p_1 = 4300 \log_{10}(h) - 3800$. d_1 and p_1 are parameters specifically defined for urban macrocellular environments. d_{2D}^l is the horizontal distance between the UAV and BS l. According to [8], the path loss with LoS and non-LoS channels are calculated as:

$$\begin{aligned} P_{NL} &= 15 + (46 - 7\log_{10}(h_l))\log_{10}\left(d_{\text{3D}}^l\right) + 20\log_{10}(f_c), \\ P_L &= 28 + 22\log_{10}\left(d_{\text{3D}}^l\right) + 20\log_{10}(f_c), \end{aligned} \tag{2}$$

where d_{3D}^{l} is the 3D distance between the UAV and BS l. f_c is the carrier frequency in GHz. h_l is the altitude of BS l.

B. Modeling of KPIs

Suppose a UAV is deployed within the service area shown in Figure 1. The UAV is required to reliably transmit data to the serving BS with little interference towards the adjacent BSs while ensuring its performance, i.e., delay and handover numbers. To evaluate whether the UAV fulfills the requirements, some KPIs are defined as follows.

1) Transmission Delay: At timestep t, $D\left(t\right)$ is used to characterize the transmission delay of the UAV. $D\left(t\right)$ depends on the data stored in the buffer denoted by $q\left(t\right)$ bits and the transmission rate $r\left(t\right)$ of the UAV. Two kinds of packets are stored in the buffer: data packets with the size of F_d bits and control packets with F_c bits. Data packets are generated following a Poisson process with the average generation rate λ. Control packets are generated by the UAV when performing handover. The UAV's stored data $q\left(t\right)$ can be modeled as

$$q(t)=q(t-1)+\sum_{n=1}^{+\infty} p(M=n|\lambda)\cdot MF_d+I(t)\cdot F_c-r(t)\cdot|t_s|, \quad (3)$$

where $p\left(M=n|\lambda\right)=\frac{e^{-\lambda}.\lambda^n}{n!}$ is the probability of generating M data packets within timestep t. $I(t)$ is the handover indicator, which equals to 1 if a handover is performed at timestep t, and 0 otherwise. Since the buffer size is limited, the maximum data volume can be stored in the UAV is q_{max}, beyond which the data will be dropped.

The coherence time of DA2G channels is assumed to be longer than $|t_s|$. Thus, the transmission rate $r\left(t\right)$ could be derived from Shannon–Hartley theorem as

$$r\left(t\right)=W_s|\mathbf{B}_k(t)|\log_2\left(1+\Gamma(t)\right), \quad (4)$$

where W_s is the frequency bandwidth of an RRB, $\mathbf{B}_k(t)$ is the set of RRBs allocated to the UAV from serving BS $k(t)$ at timestep t. $\Gamma(t)$ is the received signal-to-noise ratio (SNR) at the serving BS $k(t)$, which is

$$\Gamma(t)=P\left(\sum_{b\in\mathbf{B}_k(t)}\frac{1}{\alpha_b\left(t\right)}\right)^{-1}, \quad (5)$$

where $\alpha_b\left(t\right)=|H_b\left(t\right)|^2/\left(N_0+W_s\cdot I_b\left(t\right)\right)$. N_0 is the noise power. $I_b\left(t\right)$ is the power density of the interference over the sub-carrier of b-th RRB. $H_b\left(t\right)$ is the transfer function over the channel occupied by the sub-carrier of b-th RRB as in [23]. P is the transmission power of the UAV. The transmission delay of the UAV can be calculated as $D\left(t\right)=q\left(t\right)/r\left(t\right)$.

2) Up-link Interference: At timestep t, the up-link interference from the UAV to the adjacent BS l is denoted as

$$\mathcal{I}_l\left(t\right)=P_{dB}-P_X\left(t\right)+G_{tx}+G_{rx}, \quad (6)$$

where P_{dB} is the transmit power of the UAV in dB. $X\in\{L,NL\}$ represents the LoS and non-LoS conditions, respectively, and P_X is the path loss derived from (2). G_{tx} is the transmit antenna gain, and G_{rx} is the receive antenna gain. The sum of the up-link interference received by $L-1$ adjacent co-channel BSs can be calculated as $\sum_{l=1,l\neq k(t)}^{L}\mathcal{I}_l\left(t\right)$.

3) Handover: The last KPI is the handover performed by the UAV at timestep t, denoted by $I\left(t\right)$. $I\left(t\right)$ takes a value of 1 if a handover is processed, and 0 otherwise.

III. Problem Formulation

In this section, the optimization problem is formulated based on the above three KPIs. We assume the UAV flies at a constant altitude h with a constant velocity v, and the coordinate of it within timestep t is defined by $\{x\left(t\right),y\left(t\right)\}$. According to the previous description, UAV is assumed to complete the task over T timesteps. The trajectory design and handover management for the UAV are determined by the decisions made at each timestep t, where $t\in\{1,2,...,T\}$. The decisions consist of the moving direction, the selected serving BS $k(t)$, and the allocated RRBs $\mathbf{B}_k\left(t\right)$. Based on the proposed KPIs, the objective function is defined as,

$$\omega_t=\alpha_1\cdot D\left(t\right)+\alpha_2\cdot\sum_{l=1,l\neq k(t)}^{L}\mathcal{I}_l\left(t\right)+\alpha_3\cdot I(t), \quad (7)$$

where $\alpha_i\in[0,1]$ for $i\in\{1,2,3\}$ represents the scaling coefficient of the three KPIs, respectively. The sum of α_i equals 1, i.e., $\sum_{i=1}^{3}\alpha_i=1$, which is to balance the impact of different KPIs. The optimization problem is stated as follows:

$$\min_{(x(t),y(t),k(t),\mathbf{B}_k(t),T)}\sum_{t=1}^{T}\omega_t, \quad (8)$$

subject to:

$$\begin{aligned}
&(\mathbf{C1}) && \Gamma(t)\geq\gamma_{min},\ \forall t;\\
&(\mathbf{C2}) && \sum_{l=1}^{L}\mathbb{1}\left(l=k(t)\right)=1,\ \forall t;\\
&(\mathbf{C3}) && q\left(t\right)\leq q_{max},\ \forall t;\\
&(\mathbf{C4}) && x\left(t\right)\leq x_m,\ y\left(t\right)\leq y_m,\ \forall t;\\
&(\mathbf{C5}) && |\mathbf{B}_k(t)|\leq N_b-N_{k(t)}^{t},\ \forall t,
\end{aligned}$$

where γ_{min} is the minimum received SNR threshold. $(\mathbf{C1})$ guarantees that the UAV would not lose wireless connectivity during the movement. The function $\mathbb{1}\left(l=k(t)\right)$ is an indicator function that takes on the value 1 if the input $l=k(t)$ is true, and 0 otherwise. $(\mathbf{C2})$ guarantees that the UAV would connect with exactly one BS at each timestep. $(\mathbf{C3})$ guarantees that no data would be dropped. The variables x_m and y_m represent the maximum distance that the UAV can move in the x-axis and y-axis directions, respectively. $(\mathbf{C4})$ guarantees that the UAV moves within the service area. $(\mathbf{C5})$ indicates the maximum number of RRBs that could be allocated to the UAV.

The formulated optimization problem is challenging as the decisions are coupled to each other in time. Furthermore, since the number of terrestrial UE is a time-correlated random variable, the number of RRBs could be allocated to the UAV after fulfilling the terrestrial UE's requirement is also time-correlated. The optimal global solution requires the future knowledge of the number of terrestrial UE associated with each BS and the number of data packets expected to be generated by the UAV. Therefore, the optimization problem is complex due to the non-convexity of the objective function, its temporal coupling, and the presence of integer constraints, as well as uncertainty regarding the numbers of available RRBs and generated data packets. In order to decouple the temporal dependence of (8) and cope with the dynamic characteristics of available RRBs and generated data packets, we transform this optimization problem into a DRL problem and solve it with D3QN algorithm by considering the long-term benefits.

IV. Proposed Solution

In this section, we transform the constraints of the system model into the action space and state space and map (7) into the reward function of the D3QN algorithm. The optimization of (8) is completed by maximizing the sum of rewards during the movement of the UAV.

The set of environmental states that the ML agent could observe is called state space $\mathcal{S}$, consisting of all possible realizations of the state $S(t)$ observed at timestep t. The state $S(t)$ is composed of the coordinate of the UAV, the serving BS, the allocated RRBs to the UAV, the set of numbers of terrestrial UE and the UAV's stored data, respectively. Specifically, $S(t) = [x(t),\ y(t),\ k(t),\ \mathbf{B}_k(t),\ \mathcal{N}(t),\ q(t)]$, where $\mathcal{N}(t) = \{N_1^t, N_2^t, ..., N_L^t\}$ is the set representing the numbers of terrestrial UE being served by different BSs operating in the service area. To limit the size of the state space, we partition the service area into several square sub-areas measuring $v|t_s| \times v|t_s|$ m^2. When the UAV is inside one of the sub-areas, its coordinate is given by the coordinate of the sub-area's center. In addition, we assume that the channel conditions for the UAV remain unchanged within the same sub-area. The set of actions available to the ML agent is called action space $\mathcal{A}$. We present the action taken by the UAV at timestep $t{-}1$ in the form of a vector $A(t{-}1)$, consisting of the UAV's moving direction, the serving BS at the next timestep and the RRBs requested to be allocated by the UAV at the next timestep. Specifically, $A(t{-}1) = [\zeta(t{-}1),\ k(t),\ \mathbf{B}_k(t)]$, where $\zeta \in \mathbb{D}$ and $\mathbb{D}$ represents the possible moving directions of the UAV, which includes left, right and forward.

The reward function is defined by modifying the objective function (7) and the constraints in the optimization problem. The reward $R(t)$ obtained by the ML agent at timestep t is composed of the weighted sum of the rewards for different KPIs, i.e., R_D, R_I and R_H, the penalties P_x for $x \in \{1, 2\}$, and the bonus B. Specifically, $R_D = (1 + \beta_1 D(t))^{-1}$, $R_I = (1 + \beta_2 \sum_{l=1, l \neq k(t)}^{L} \mathcal{I}_l\,(t))^{-1}$, and $R_H = (1 + \beta_3 I\,(t))^{-1}$. Thus, the reward $R(t)$ is formulated as

$$\begin{aligned} R\,(t) = & \alpha_1 \cdot R_D + \alpha_2 \cdot R_I + \alpha_3 \cdot R_H \\ & + \mathbb{1}((x(t) \notin [0, x_m]) \vee (y(t) \notin [0, y_m])) \cdot P_1 \\ & + \mathbb{1}\,(\Gamma(t) < \gamma_{min}) \cdot P_2 \\ & + \mathbb{1}(d(t{-}1) > d(t)) \cdot B, \end{aligned} \tag{9}$$

where $d(t) = \left((x(t) - x_e)^2 + (y(t) - y_e)^2\right)^{1/2}$ is the distance between the UAV and the end point $\{x_e, y_e\}$ at timestep t, β_x for $x \in \{1, 2, 3\}$ is used to balance the relative magnitudes of different KPIs. Indicator function $\mathbb{1}$ takes a value of 1 if the input is true, and 0 otherwise. The rewards for different KPIs are normalized to the $[0, 1]$ interval. In addition to the rewards related with KPIs, we give additional penalties and bonus for some specific situations to ensure the UAV can reach the target. The penalty P_1 is implemented to prevent the UAV from exceeding the borders of the service area, the penalty P_2 applies when the UAV loses wireless connectivity, and the bonus B is granted for the UAV moving closer to the target.

Since our objective is to jointly optimize the KPIs during the task, $\sum_{t=1}^{T} R\,(t)$ should be considered. As the actions of the UAV are related to each other in sequence, greedily maximizing the reward at each timestep t would not be the globally optimal solution. Thus, we propose D3QN algorithm to achieve joint optimization. D3QN algorithm is the improved version of DQN algorithm, which adds the evaluation of the states to accelerate convergence and eliminates the overestimation problem by decoupling action selection and Q-value calculation. D3QN algorithm aims to maximize long-term rewards, and is able to predict future states according to current observation, which fits well with our proposed optimization problem. By discounting the expected future reward to the current state, the D3QN agent can decouple the UAV's actions at different timesteps and choose the one with the highest expected reward, thereby optimizing the reward globally.

The proposed D3QN algorithm is described in detail as follows. First, we initialize the parameters θ and θ^- of the evaluate network and the target network, respectively. Afterwards, Algorithm 1 is used to train these two neural networks. Specifically, we utilize the stochastic gradient descent (SGD) method to update the evaluate network, which is to compute the gradient of the expected reward with respect to the network parameters based on randomly selected training records. In addition, we assign the network parameters of the evaluate network to the target network at intervals to update the target network. When the training is complete, the evaluate network is saved and loaded into the ML agent to instruct the UAV to complete the task. The testing is processed as Algorithm 2.

V. Simulation Results

We consider a rectangular urban service area measuring 500×500 m^2, where 4 BSs are located at its corners, serving both air and terrestrial UE. A UAV is deployed in this area as the air UE as shown in Figure 1. An ML controller performs the trajectory design and handover management for the UAV according to the D3QN algorithm. To evaluate the performance of the D3QN algorithm, we consider a greedy policy as the benchmark scheme. In the greedy policy, the UAV moves along the shortest straight line to reach the end point, and all the available RRBs from the serving BS are allocated to the UAV during its movement. In addition, the decision of performing a handover is based on RSS with a margin of 7 dB, and a time to trigger (TTT) of $|t_s|$ is employed. Specifically, if the power received by the UAV from an adjacent BS is 7 dB stronger than that of the serving BS for a duration of $|t_s|$, the UAV would perform a handover. After training the D3QN agent, we conduct 5000 tests to

Algorithm 1 Training Process of D3QN Algorithm

```
Initialize: Evaluate network parameters $\theta$, target network parameters $\theta^{-}$, reply buffer $B$, training batch size $B_b$, network replacement frequency $f_r$
for episode $e \in \{1, 2, ..., E\}$ do
  Initialize sequence $s_1 = \{x_1\}$ as the initial state.
  Standardize the initial state $\phi_1 = \phi\{s_1\}$.
  for timestep $t \in \{1, 2, ..., T\}$ do
    Set action $a_t = \arg\max_a Q(\phi_t, a; \theta)$.
    Obtain and standardize next state $\phi_{t+1} = \phi\{s_{t+1}\}$.
    Obtain reward $r_t$ according to (9).
    Store tuple $(\phi_t, a_t, r_t, \phi_{t+1})$ in $B$.
    Random select mini-batch $\mathcal{J}$ from $B$.
    for every tuple $(\phi_j, a_j, r_j, \phi_{j+1})$ in $\mathcal{J}$ do
      if UAV reaches the target at timestep $j + 1$ then
        $y_j = r_j$.
      else
        Obtain $a_{max} = \arg\max_a Q(\phi_{j+1}, a; \theta)$.
        $y_j = r_j + \gamma Q(\phi_{j+1}, a_{max}; \theta^{-})$.
      end if
      Gradient descent with $\|y_j - Q(\phi_j, a_j; \theta)\|^2$.
      if $\mod (t, f_r) == 0$ then
        $\theta^{-} = \theta$.
      end if
    end for
  end for
end for
```

Algorithm 2 Testing Process of D3QN Algorithm

```
Initialize: As in Algorithm 1.
if UAV does not reach the target at timestep $t$ then
  Obtain and standardize current state $\phi_t = \phi\{s_t\}$.
  Set action $a_t = \arg\max_a Q(\phi_t, a; \theta)$.
  Obtain reward $r_t$ according to (9).
end if
```

evaluate its performance with different scaling coefficients. We consider four different scaling coefficient combinations, three of which represent policies that focus on prioritizing the optimization of a specific KPI over others while still optimizing all KPIs. These policies include the "min-delay policy" which prioritizes optimization of the transmission delay, the "min-interference policy" which prioritizes reducing up-link interference, and the "min-handover policy" which gives priority to the reduction of the number of handovers. The fourth and final policy is the "joint optimization policy", which aims to optimize all KPIs in a balance without any specific prioritization. The parameter settings of the simulation are shown in Table I.

The joint optimization policy is proposed to address the issue of high up-link interference and many redundant handovers for the UAV. Thus, we assign a relatively small value to the scaling coefficient of the delay-related KPI and analyze the dependence of R_I and R_H in (9) on their respective coefficients. The results are shown in Figure 2, we set $\alpha_1 = 0.2$, and investigate the impact of changing α_2 while maintaining $\alpha_2 + \alpha_3 = 0.8$. It should be emphasized that the value of α_1 is not immutable, we choose a small value to better evaluate the performance improvement on UAV's up-link interference and handover numbers. However, If operators aim to further reduce the transmission delay of the UAV, they can set a higher value for α_1. With the min-delay policy, we verify that increasing the value of α_1 can enable the D3QN algorithm to achieve superior delay performance compared to the benchmark scheme, indicating the D3QN algorithm can be applied to scenarios with stringent delay requirements.

TABLE I
PARAMETERS FOR THE SIMULATION.

Parameters	**Values**
Service area	500 m × 500 m
Available RRBs and height of UAVs	Up to 4×180 kHz, 50 m
Numbers and antenna height of BSs	4, 25 m
Packet arrival rate and size	50 Hz, 2 kbits
Handover control packet size	4×1200 bits
The length of a timestep $\lvert t_s \rvert$	100 milliseconds
The balance coefficients β_x	$\beta_1 = 10^5$, $\beta_2 = 10^{11}$, $\beta_3 = 10$
Min-delay policy	$\alpha_1 = 80\%$, $\alpha_2 = \alpha_3 = 10\%$
Min-interference policy	$\alpha_2 = 80\%$, $\alpha_1 = \alpha_3 = 10\%$
Min-handover policy	$\alpha_3 = 80\%$, $\alpha_1 = \alpha_2 = 10\%$
Joint optimization policy	$\alpha_1 = 20\%$, $\alpha_2 = 50\%$, $\alpha_3 = 30\%$

Figure 2 shows the average up-link interference and handover numbers obtained from 5000 tests for different values of α_2. The up-link interference and handover numbers display inverse trends with the increase of α_2, indicating a trade-off that optimizing one factor comes with the expense of the other. As the α_2 increases, the up-link interference of the UAV shows a significant decreasing trend. For the case of $\alpha_2 = 0.5$, the D3QN algorithm achieves similar interference performance as the benchmark scheme, while reducing the handover number by about 90%. However, if α_2 is higher than 0.5, the reduction of up-link interference is relatively limited, and the handover number exhibits a significant upward trend. Therefore, we adopt the setting of $\alpha_2 = 0.5$ for the joint optimization policy.

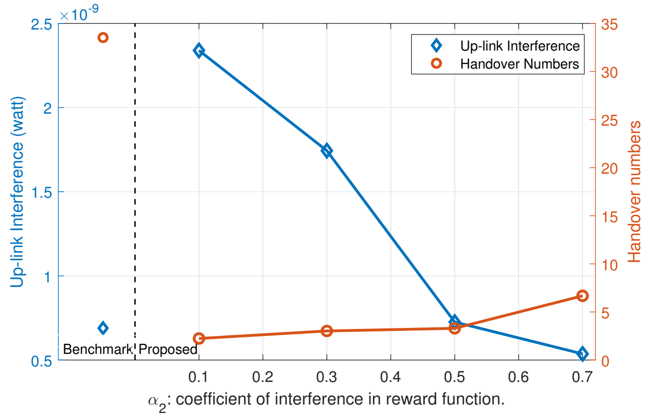

Fig. 2. Performance evaluation for different coefficients of interference.

We give an example of trajectory design based on the D3QN algorithm with different policies and compare it with the benchmark scheme in Figure 3. The markers placed along the trajectories indicate the positions where the UAV performs handovers. The trajectories designed by the D3QN algorithm for different policies exhibit similar characteristics, as they guide the UAV to bypass the central region while approaching the target, thereby avoiding frequent handovers. Only a limited number of handovers occur within the overlapping coverage areas of different BSs, such as the central region of the service area, which illustrates that the D3QN algorithm effectively eliminates a large number of unnecessary handovers for the UAV through trajectory design and handover management. However, when executing the greedy policy, the UAV moves directly through the central region to reach its target. Since the central region is around the boundaries of the cells served by different BSs, the UAV frequently switches to the BS with the highest RSS when passing through it, resulting in many redundant handovers.

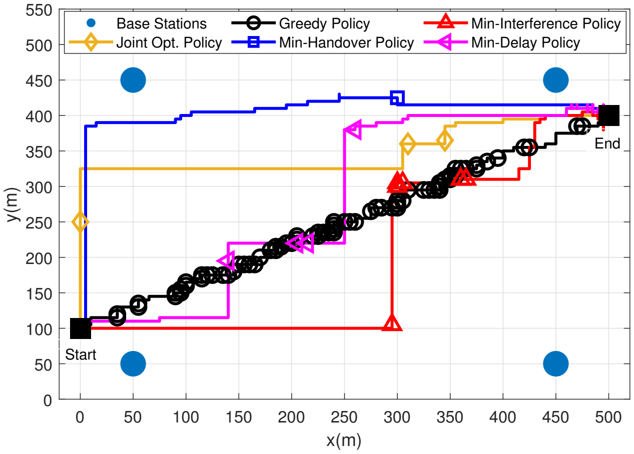

Fig. 3. Different trajectories designed by D3QN algorithm and greedy policy.

From Figure 4 to Figure 6, we show the performances of the D3QN algorithm on three KPIs under different policy designs and compare them with the benchmark scheme. In addition, the performance on various KPIs of the D3QN algorithm without trajectory design is also evaluated by implementing the joint optimization policy without trajectory design (TD). Specifically, the same coefficients as the joint optimization policy are assigned to this policy. However, the UAV should move from the start point to the end point along the shortest straight path and the UAV's performance can only be improved through handover management.

Figure 4 shows the performance of the D3QN algorithm in optimizing the number of handovers. The figure illustrates the cumulative distribution function (CDF) of the number of handovers performed by the UAV during the entire path. For the greedy policy, the UAV passes through the central region and always connects to the BS with the highest RSS to obtain the highest SNR, resulting in the ping-pong effect. Thus, the number of handovers varies between 20 to 50, and many of the handovers are unnecessary.

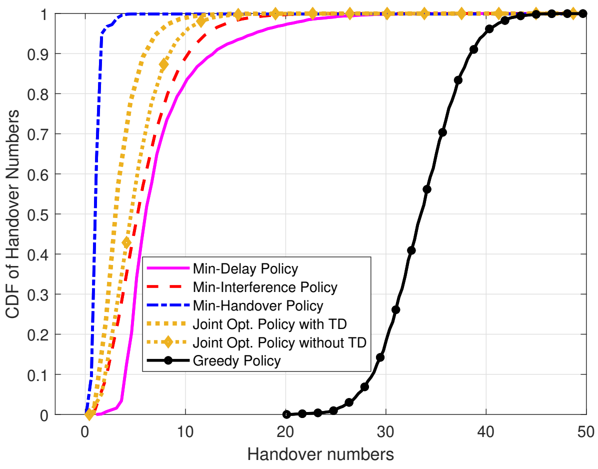

Fig. 4. The CDF of handover numbers.

However, the D3QN algorithm can eliminate the ping-pong effect and significantly reduce the number of handovers. The min-handover policy focuses on reducing the number of handovers, with over 95% probability that only one handover is required. The handover is initiated only to avoid the loss of wireless connectivity. Compared with the min-handover policy, applying the joint optimization policy with TD results in a slightly higher number of handovers. However, it still achieves an average reduction of about 90% in handovers compared to the greedy policy, as illustrated by the median value of the handover numbers in Figure 4. Trajectory design helps the joint optimization policy to reduce the number of handovers, which can be illustrated by comparing the joint optimization policy with and without TD. In addition, the min-delay and min-interference policies control the UAV to perform slightly more handovers than the joint optimization policy with TD because the UAV can only obtain limited rewards from reducing handover numbers. However, these policies can still maintain the number of handovers less than 10 in more than 80% of the cases. Although the transmission delay and up-link interference are affected by the amount of randomly generated data from the UAV and the randomness of terrestrial UE, respectively, these two policies still significantly reduce the handover numbers compared to the greedy policy.

Figure 5 illustrates the D3QN algorithm's performance in optimizing up-link interference. The x-axis represents the average up-link interference caused by the UAV during the entire movement, while the y-axis shows the corresponding cumulative probability. When applying the greedy policy, the UAV effectively reduces the up-link interference to adjacent BSs by frequently performing handovers. Specifically, signals propagated by the UAV to different BSs will experience path losses of different magnitudes. The UAV obtains high RSS by frequently switching to the BS with the lowest path loss, which leads to the fact that the wireless propagation of the UAV to adjacent BSs will experience high path loss, thereby reducing the up-link interference towards adjacent BSs. However, despite the achievable high RSS, the requirement for all available RRBs can produce up-link interference over a wider bandwidth compared to policies based on the D3QN algorithm.

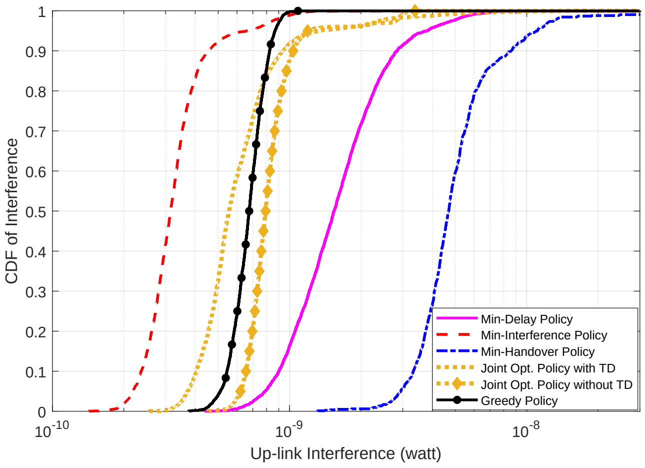

Fig. 5. The CDF of up-link interference.

The min-interference policy achieves a significant decrease in the up-link interference caused by the UAV, with a reduction of approximately 53% or 3.3 dB compared with the greedy policy as illustrated from the median values of the up-link interference in Figure 5. The min-interference policy mitigates up-link interference by performing trajectory design and reducing the number of allocated RRBs. Specifically, the UAV reduces up-link interference by performing trajectory design to avoid the LoS connectivity with non-associated BSs and by requiring a rather small number of allocated RRBs. The performance of the joint optimization policy with TD is more balanced, since it allocates more RRBs for the UAV to increase the throughput and maintain low delay, but slightly sacrifices the performance of up-link interference. Thus, the joint optimization policy with TD performs slightly worse in up-link interference when compared with the min-interference policy, but it can still achieve a reduction of approximately 18% or 0.84 dB in comparison with the greedy policy as illustrated from the median values. However, the joint optimization policy without TD increases up-link interference by about 16% or 0.65 dB compared with the greedy policy as illustrated from the median values, emphasizing the importance of trajectory design in interference reduction. The min-delay policy and min-handover policy cause higher up-link interference than the greedy policy since the reward function of the D3QN algorithm corresponding to the two policies assigns a small coefficient to up-link interference, making the UAV obtain limited rewards from reducing interference. The min-handover policy results in higher interference than the min-delay policy due to the fewer handovers performed by the UAV.

The performance of the D3QN algorithm in optimizing transmission delay is shown in Figure 6. The x-axis represents the transmission delay of the UAV at each timestep, while the y-axis represents the corresponding cumulative probability. The greedy policy employed by the UAV prioritizes the BS with the best channel condition to achieve a high transmission rate. However, this approach leads to more handovers, generating a large number of control packets that increase the transmission burden. As a result of the heavy transmission burden, it is hard for the greedy policy to achieve a low transmission delay even with a high transmission rate.

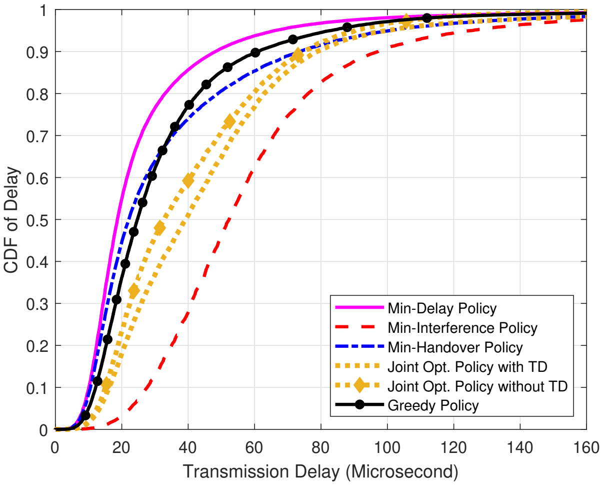

Fig. 6. The CDF of transmission delay.

When the min-handover policy is applied, the behavior of the UAV differs significantly from that of the greedy policy by performing only one handover in most cases. Although this policy may not achieve as high throughput as the greedy policy, it prevents the UAV from being burdened in transmission due to excessive control packets from frequent handovers. Consequently, the UAV exhibits a performance of transmission delay comparable to that achieved by the greedy policy. However, the performance of the two mentioned policies is suboptimal. The min-delay policy optimizes transmission delay through trajectory design and proper handovers, achieving a median transmission delay of 19 microseconds, which is 6 microseconds less than that of the greedy policy.

The joint optimization policy with TD displays a decline in transmission delay performance, resulting in a median transmission delay that is 15 microseconds higher than that of the greedy policy. The performance degradation is because the policy focuses on addressing high interference and redundant handover, and it assigns a small value to the scaling coefficient of transmission delay. The joint optimization policy without TD performs slightly better than the one with TD in transmission delay, as the UAV would not move closer to the edge of the service area, resulting in higher throughput. Specifically, if the UAV is located at the edge and cannot connect to the nearest BS due to the non-LoS connectivity, the distance between the UAV and the serving BS would be farther compared with the situation where the UAV is in the central region, which leads to lower throughput and higher transmission delay. Finally, the min-interference policy exhibits the highest delay as it pays less attention to optimizing for transmission delay.

To comprehensively evaluate the performance for different policies, we tabulate the median values of the performance concerning different KPIs and present them in Table II. All of the policies prioritizing the optimization of a specific KPI over others, i.e., the min-delay policy, the min-handover policy, and the min-interference policy, outperform the benchmark scheme in the relative KPI. The joint optimization policy with TD is the policy with balanced performance, which performs better than the benchmark scheme in terms of up-link interference and handover numbers, indicating its superiority in addressing high interference and redundant handovers. However, the joint optimization policy without TD is inferior to the corresponding policy with TD in terms of up-link interference and handover numbers, highlighting the effectiveness of trajectory design in improving the performance of the UAV.

TABLE II
MEDIAN OF THE PERFORMANCE FOR DIFFERENT POLICIES.

Policy	**Delay (microsecond)**	**Interference (10^{-10} watt)**	**Handover numbers**
Min-interference	52.3	3.2	5.2
Min-handover	22.0	46.9	1.0
Min-delay	18.7	15.5	6.0
Joint opt. with TD	39.4	5.6	3.0
Joint opt. without TD	32.8	7.9	4.5
Greedy algorithm	24.8	6.8	33.5

VI. CONCLUSION

The main challenges addressed in this paper include the up-link interference from UAVs towards terrestrial BSs and the redundant handovers performed by UAVs while moving. To address these issues, the proposed D3QN algorithm designs different rewards for the states and actions of the UAV, thereby transferring the trajectory design and handover management into policies that pursue high rewards. The D3QN algorithm discounts the expected future rewards into the current available rewards related to different actions, thereby decoupling the temporal correlation of the UAV's action selections and achieving global optimization. The simulation results demonstrate the superiority of our proposed algorithm. On the one hand, by prioritizing a single KPI over the others, the proposed algorithm outperforms the benchmark scheme in the prioritized KPI at the expense of others. Specifically, the proposed algorithm can eliminate up to 95% of the handovers, reduce 25% of the transmission delay, or reduce the up-link interference by 53%, respectively. On the other hand, we investigate the trade-off between the up-link interference and handover numbers, and design a policy with a balanced performance called the joint optimization policy with TD to help the UAV eliminate unnecessary handovers while reducing uplink interference. This policy achieves joint optimization for eliminating about 90% of the handovers and 18% of the up-link interference at the cost of a delay increment of 15 microseconds, when compared with the benchmark scheme. Finally, we also investigate the performance gains of trajectory design. For the joint optimization policies, the policy considering trajectory design performs better than the one without trajectory design. On average, the former reduces the up-link interference of the latter by 29% and handovers by 33%, albeit at the cost of a delay increment of 6 microseconds.

REFERENCES

[1] A. Baltaci *et al.*, "A survey of wireless networks for future aerial communications (facom)," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2833–2884, 2021.

[2] S. Zhang and N. Ansari, "Latency aware 3d placement and user association in drone-assisted heterogeneous networks with fso-based backhaul," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11991–12000, 2021.

[3] Y. Zeng, J. Lyu, and R. Zhang, "Cellular-connected uav: Potential, challenges, and promising technologies," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 120–127, 2019.

[4] E. Dinc, M. Vondra, and C. Cavdar, "Total cost of ownership optimization for direct air-to-ground communication networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10157–10172, 2021.

[5] E. Dinc, M. Vondra, and C. Cavdar, "Multi-user beamforming and ground station deployment for 5g direct air-to-ground communication," *IEEE Global Communications Conference*, pp. 1–7, 2017.

[6] I. A. Meer, M. Ozger, and C. Cavdar, "On the localization of unmanned aerial vehicles with cellular networks," *IEEE Wireless Communications and Networking Conference*, pp. 1–6, 2020.

[7] L. Sundqvist, "Cellular Controlled Drone Experiment: Evaluation of Network Requirements," master's thesis, Aalto University. School of Electrical Engineering, 2015.

[8] 3GPP TR 36.777, "Enhanced lte support for aerial vehicles," 2018.

[9] M. M. Azari, F. Rosas, A. Chiumento, and S. Pollin, "Coexistence of terrestrial and aerial users in cellular networks," *IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2017.

[10] A. Fakhreddine *et al.*, "Handover challenges for cellular-connected drones," in *Proceedings of the 5th Workshop on Micro Aerial Vehicle Networks, Systems, and Applications*, pp. 9–14, 2019.

[11] J. Chen and D. Gesbert, "Optimal positioning of flying relays for wireless networks: A los map approach," *IEEE International Conference on Communications*, pp. 1–6, 2017.

[12] R. Gangula *et al.*, "A landing spot approach for enhancing the performance of uav-aided wireless networks," *IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2018.

[13] H. Huang *et al.*, "Deep reinforcement learning for uav navigation through massive mimo technique," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1117–1121, 2020.

[14] U. Challita, W. Saad, and C. Bettstetter, "Interference management for cellular-connected uavs: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2125–2140, 2019.

[15] Y. Chen *et al.*, "A deep learning approach to efficient drone mobility support," *Proceedings of the 2nd ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and Beyond*, pp. 67–72, 2020.

[16] A. Azari, F. Ghavimi, M. Ozger, R. Jantti, and C. Cavdar, "Machine learning assisted handover and resource management for cellular connected drones," *IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–7, 2020.

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," 2013.

[18] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," 2015.

[19] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Freitas, "Dueling network architectures for deep reinforcement learning," 2015.

[20] F. Salehi, M. Ozger, N. Neda, and C. Cavdar, "Ultra-reliable low-latency communication for aerial vehicles via multi-connectivity," *Joint European Conference on Networks and Communications & 6G Summit*, pp. 166–171, 2022.

[21] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on uavs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[22] M. Ozger, M. Vondra, and C. Cavdar, "Towards beyond visual line of sight piloting of uavs with ultra reliable low latency communication," *IEEE Global Communications Conference*, pp. 1–6, 2018.

[23] M. Kalil, A. Shami, A. Al-Dweik, and S. Muhaidat, "Low-complexity power-efficient schedulers for lte uplink with delay-sensitive traffic," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4551–4564, 2015.