

Mobility Management for Cellular-Connected UAVs: A Learning-Based Approach

Md Moin Uddin Chowdhury*, Walid Saad[†], and Ismail Güvenc*

*Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA

[†]Wireless@VT, Electrical and Computer Engineering Department, Virginia Tech, VA, USA

Email: {mchowdh, iguvenc}@ncsu.edu, walids@vt.edu

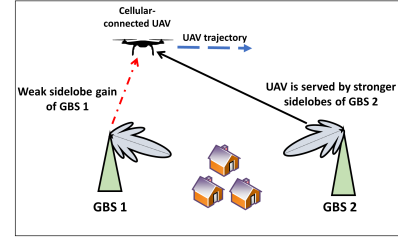
Abstract—The pervasiveness of the wireless cellular network can be a key enabler for the deployment of autonomous unmanned aerial vehicles (UAVs) in beyond visual line of sight scenarios without human control. However, traditional cellular networks are optimized for ground user equipment (GUE) such as smartphones which makes providing connectivity to flying UAVs very challenging. Moreover, ensuring better connectivity to a moving cellular-connected UAV is notoriously difficult due to the complex air-to-ground path loss model. In this paper, a novel mechanism is proposed to ensure robust wireless connectivity and mobility support for cellular-connected UAVs by tuning the downtilt (DT) angles of all the ground base stations (GBSs). By leveraging tools from reinforcement learning (RL), DT angles are dynamically adjusted by using a model-free RL algorithm. The goal is to provide efficient mobility support in the sky by maximizing the received signal quality at the UAV while also maintaining good throughput performance of the ground users. Simulation results show that the proposed RL-based mobility management (MM) technique can reduce the number of handovers while maintaining the performance goals, compared to the baseline MM scheme in which the network always keeps the DT angle fixed.

Index Terms—3GPP, antenna radiation, mobility management, reinforcement learning, trajectory, UAV.

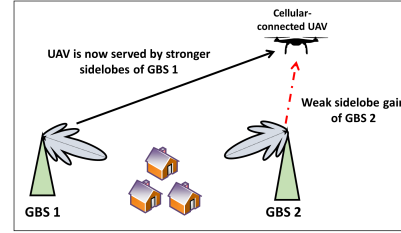
I. INTRODUCTION

The use of unmanned aerial vehicles (UAVs) in both civilian and commercial applications is a promising technology for next-generation cellular networks. Thanks to their ease of deployment, high flexibility, and the ability to increase network capacity, UAVs can be deployed in surveillance, remote sensing, package delivery, and disaster-relief applications. Hence, significant efforts in both academia and industry have been devoted to different aspects of UAV applications. However, most of the above-mentioned UAV cases are operated by trained pilots.

To fully reap the benefits of UAV deployment, beyond visual line of sight operations are of critical importance where UAVs acting as aerial users, can maintain communication with the ground base stations (GBSs) for command and control (C&C) purposes in the downlink (DL) [1]. UAVs flying in the sky may be served by the sidelobes of base station antennas which provide lower antenna gains [1], [2]. This poses major challenges on the mobility management (MM) for the cellular-connected UAVs based on reference signal received power (RSRP). The GBS providing the highest RSRP might be located far away from the UAV. An example is shown in Fig. 1, where a cellular-connected UAV is flying over a rural



(a) Scenario: 1



(b) Scenario: 2

Fig. 1. HO scenario of a cellular-connected UAV moving towards horizontal direction. (a) UAV is associated with GBS 2 due to its higher side-lobe gain than GBS 1. (b) After moving forward, the UAV is now associated with GBS 1 due to its higher sidelobe gain.

environment. This type of patchy signal coverage of GBSs will result in poor mobility performance such as handover failure (HOF), radio link failure, as well as unnecessary handovers (HOs), called ping-pong events. Apart from these, due to the loss of the C&C signal, the UAV may collide with a commercial aircraft or even crash into a populated area which might result in hazardous events. Hence, effective MM for providing reliable communications between UAVs and GBSs is of critical importance.

MM techniques for ground user equipment (GUE) in both homogeneous and heterogeneous cellular networks have been studied extensively in the literature [3]–[5]. However, the research in MM for cellular-connected UAVs is still in its infancy. Motivated by this, considering an interference-limited DL cellular network, the main contribution of this paper is a novel a reinforcement learning (RL) based MM technique that can ensure better connectivity at the UAV while also maintaining a reasonable throughput performance for existing GUE. The proposed approach exploits Q-learning [6] for tuning the downtilt (DT) angles of GBSs to meet these goals. In this regard, we assume that the UAV's trajectory is known beforehand and the network can control the RSRP values at different discrete waypoints of the predefined UAV trajectory

by changing the DT angles. We also consider the presence of correlated shadow fading [7] for the predetermined UAV trajectory. We introduce a reward function that takes into account both RSRP at the UAV and the capacity of the GUE, and we study the performance of different weight combinations associated with both of these performance criteria. Our simulations demonstrate a tradeoff between the RSRP quality at the UAV and the GUE capacity. Moreover, in a scenario, where the network considers maximizing the RSRP values, the number of HOs can be reduced by 40%, compared to a scheme where the DT angle is remained fixed.

The rest of this paper is outlined as follows. In Section II, we provide a literature review pertinent to MM of cellular-connected UAVs. Section III describes the system model. We discuss the proposed MM technique in Section IV. Related simulations results are presented in Section V. Finally, Section VI concludes this paper.

II. RELATED WORKS

Integrating UAVs as aerial user equipment with existing GUE has attracted attention in recent years. For instance, authors in [8] presented an analytical framework for co-existing UAV and GUE considering beamforming technique. In [9], authors provided the upper and lower bounds on the coverage probability of UAVs considering a coordinated multi-point technique. Authors in [10] investigated the problem of interference-aware optimal path planning of cellular-connected UAVs using deep RL in the uplink (UL) scenario. In [11], authors explored the effects of practical antenna configurations on the mobility of cellular-connected UAVs and showed that increasing the number of antenna elements can increase the number of HOs for vertically-mobile UAVs.

In [12] and [13], real-world industrial experiments were conducted to test the feasibility of integrating UAVs in existing cellular networks. The Third Generation Partnership Project (3GPP) also studied the potential of integrating UAVs as aerial UEs in existing cellular networks and pointed out the challenges in providing reliable mobility support [14]. According to [15], downlink (DL) performance and UL interference can be mitigated by introducing antenna selection or beamforming capabilities at the UAVs. By conducting extensive 3GPP compliant simulations, in [16], the authors showed that existing cellular networks will be able to support a small number of aerial user equipment (UE) with good mobility support.

In recent work, the authors explored the RL algorithm to maximize the received signal quality at a cellular-connected UAV while minimizing the number of HO [17]. A novel HO performance optimization algorithm by tuning the values of the 3GPP specified HO parameters in an automated manner was introduced in [18]. An RL-based HO optimization scheme for ground UEs in a 5G cellular network was proposed in [19].

However, none of these prior works [8]–[15], [17], [19] studied the problem of concurrently maintaining good RSRP quality for the UAV and high throughput for GUE while considering practical antenna pattern of the GBS. It is worth noting that the network can control the received power quality of the users by changing the DT angle [20]. DT angles can be easily and remotely carried out by adjusting the relative phases of antenna elements of an antenna array in such a way

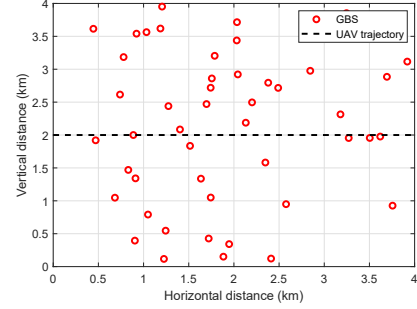


Fig. 2. Illustration of cellular network with linear mobility model.

that the radiation pattern can be downtilted uniformly in all horizontal directions [21]. To the best of our knowledge, no prior work has considered tuning DT angles dynamically for effective MM of cellular-connected UAVs.

III. SYSTEM MODEL

A. Network Model

Consider the DL of an interference-limited cellular network in which a single UAV acting as an aerial user, is flying along a two dimensional (2D) linear trajectory at a fixed height h_{uav} . The network consists of M GBSs and K static GUE with similar height h_{GUE} . We also assume that all GBSs have equal altitudes h_{GBS} and transmission power P_{GBS} . The set of the GUE can be denoted as \mathcal{K} with horizontal coordinates $\mathbf{r}_k = [x_k, y_k, z_k]^T \in \mathbb{R}^{3 \times 1}, k \in \mathcal{K}$. We define \mathcal{M} as the set of the GBSs. The GBSs have three sectors separated by 120° , while each sector is equipped with 8×1 cross-polarized antennas downtilted by the same DT angle β . We assume that the UAV and the GUE are each equipped with an omnidirectional antenna. We consider the 3GPP antenna radiation model to characterize the antenna radiation at the GBS [22].

We consider the sub-6 GHz band for the cellular network and, hence, the presence of thermal noise power at the receivers is negligible compared to the interference power. We also assume that the GBSs share a common transmission bandwidth and full buffer traffic is used in every cell. All DL transmissions are scheduled using a round-robin scheduling algorithm, and all the receivers are capable of mitigating the Doppler effect [23], [24].

B. UAV Mobility Model

A simple UAV mobility scenario is considered whereby the UAV travels along a linear trajectory (for instance, through the horizontal x-axis) with constant speed v km/h and a fixed altitude of h_{uav} . We consider the linear mobility model due to its simplicity and suitability for UAVs flying in the sky with virtually no obstacle. In Fig. 2, we provide an illustrative example of the linear UAV mobility model for $v = 120$ km/h in an area of 4×4 km² with $M = 48$.

C. Path-loss Models and GUE Capacity

The path-loss between a GBS and a UAV plays a significant role in the HO performance due to the RSRP-based cell association. For modeling the path-loss, we consider the RMa-AV-LoS channel model specified in 3GPP [14]. The instantaneous

path-loss (in dB) under a line-of-sight (LOS) scenario between GBS m and the UAV can be expressed as:

$$\begin{aligned} \xi_{m,u}^{\text{LOS}}(t) = & \max(23.9 - 1.8 \log_{10}(h_{\text{uav}}), 20) \log_{10}(d_{m,u,t}) \\ & + 20 \log_{10}\left(\frac{40\pi f_c}{3}\right) + \chi_{\text{LOS}}, \end{aligned} \quad (1)$$

where h_{uav} is between 10 m to 300 m and f_c is the carrier frequency, while $d_{m,u,t}$ represents the 3D distance between the UAV and GBS m at time t . χ_{LOS} represents the correlated shadow fading (SF) (in dB) associated with LOS scenario. We will discuss about SF in the next subsection. It is worth noting that the probability of LOS is equal to one if the UAV height falls between 40 m and 300 m [14]. Finally, we can calculate the RSRP (in dB) from GBS m at time t as $\text{RSRP}_{m,u}(t) = P_{\text{GBS}} + G_{m,u}(t) - \xi_{m,u}^{\text{LOS}}(t)$, where $G_{m,u}(t)$ is the antenna gain at the UAV from GBS m at time t , which is dependent on the UAV's location and β [1].

The path-loss (in dB) observed at GUE $k \in \mathcal{K}$ from GBS m is given by: $\xi_{k,m} = A + B \log_{10}(d_{k,m}) + C$. Here, $d_{k,m}$ is the Euclidean distance from GBS m to GUE k . A , B , and C are factors dependent on the carrier frequency f_c and antenna heights [25]. Then, the received power at GUE k from GBS m can be calculated as $S_{k,m} = \frac{P_{\text{GBS}} G_{k,m}}{10^{(\xi_{k,m}/10)}}$, where $G_{k,m}$ represents the antenna gain at GUE k from GBS m that depends on the locations of GUE k and GBS m [22]. Note that $G_{k,m}$ will be dependent on β as shown in [1]. A GUE associates with the GBS that provides the best signal-to-interference ratio (SIR). The SIR between GUE k and GBS m can be calculated as $\gamma_{k,m} = \frac{S_{k,m}}{\sum_{j \in \mathcal{M}, j \neq m} S_{k,j}}$. Assuming that, GUE k is associated with GBS m , we can express the achievable data rate per unit bandwidth (bps/Hz) of GUE k using Shannon's capacity as:

$$C_{k,m} = \frac{\log_2(1 + \gamma_{k,m})}{N_{\text{GUE}}^m}, \quad (2)$$

where N_{GUE}^m is the number of GUEs associated with GBS m . Since we are considering round-robin scheduling with full buffer traffic, the available resources will be equally divided among the GUEs associated with a particular GBS. Assuming that the set of GUEs associated with GBS m is \mathcal{K}_m , the total GUE capacity can be calculated as $C = \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}_m} C_{k,m}$.

D. Shadow Fading Model

SF also known as medium-scale fading, plays a critical role in characterizing the received power at the receiver. It is caused mainly due to the presence of large obstacles within a wireless link [26]. SF is typically modeled as an independent Gaussian random variable with zero mean and standard deviation σ . According to [14], σ (in dB) can be expressed as $\sigma = 4.2 \exp(-0.0046 h_{\text{uav}})$. However, the SF values of consecutive waypoints of a UAV trajectory might have some similarities or correlation due to high probability of LOS in the GBS-to-UAV link. Based on empirical measurements, the work in [27] first proposed the common analytical model for auto-correlation among the SF values which assumes that SF is a first-order auto-regressive process where the auto-correlation

between the SF values at two points separated by distance Δ is given by [26]:

$$R(\Delta) = \sigma^2 \rho^{\frac{\Delta}{X_c}}, \quad (3)$$

where, Δ is the distance between the two points, ρ is the correlation coefficient, and X_c is the decorrelation distance [26]. Here, we set $X_c = 100$ m and $\rho = 0.82$ [26]. We first generate independent SF values with zero mean and $\sigma = 1$ for each waypoint and then use Cholesky factorization to generate the correlated SF values from $R(\Delta)$ [28].

E. Handover Procedure

Throughout the flight duration, the UAV can measure the RSRPs from all the adjacent GBSs at subsequent measurement gaps using (1). We consider a HO mechanism that involves a HO margin (HOM) parameter, and a time-to-trigger (TTT) parameter, which is a time window which starts after the following HO condition (A3 event [29]) is fulfilled.

$$\text{RSRP}_j > \text{RSRP}_i + m_{\text{hyst}}, \quad (4)$$

where RSRP_j and RSRP_i are the RSRP measured from the serving GBS i and target GBS j , respectively and m_{hyst} is the HOM set by the network operator. The UAV does not transmit its measurement report to its current serving GBS before the TTT timer expires [4].

IV. RL-BASED MOBILITY MANAGEMENT

In this section, we will describe our proposed RL-based MM framework. The aim is to determine the optimal sequence of DT angles for each discrete waypoint of the predefined UAV trajectory for maintaining good RSRP quality at the UAV while maintaining good GUE rate performance. This problem can be solved by non-linear optimization techniques [30]. However, optimization-based techniques require the exact knowledge of the modeling parameters, which might not always be available. Moreover, even with the perfect information of all relevant parameters, such an optimization problem is NP-hard [31] and difficult to solve efficiently due to the intractability in GBS antenna radiation pattern, HO process, and channel models considered in this research.

Thus, we consider RL algorithm for solving this sequential decision problem in which an agent tries to maximize its cumulative rewards by interacting with an unknown environment through time [32]. The agent gets an immediate reward or punishment for taking any action and tries to maximize the total expected future reward. Usually, the environment is modeled as a Markov decision process (MDP) which is characterized by a tuple $(\mathcal{S}, \mathcal{A}, P, R)$, where \mathcal{S} is the set of states, \mathcal{A} is the set of actions, P is the state transition probability function, with $P(s'|s, a)$ denotes the probability of moving to the next state $s' \in \mathcal{S}$ from the current state $s \in \mathcal{S}$ after taking action $a \in \mathcal{A}$, and $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the immediate reward received by the agent. An MDP can be solved with this information to obtain the optimal policy, i.e., the action to take at each state which will maximize the expected sum of discounted future rewards.

A branch of RL known as model-free RL algorithms can learn optimal policies in finite MDPs without explicit knowledge about the environment modeling, i.e., P and R . They are typically simpler and more flexible to implement than

their model-based counterpart since the dynamics or the model of the environment is not required to be known in prior. In a typical cellular network, to obtain the perfect knowledge of the environment, a significant amount of information has to be exchanged between the network components and the core network, which is not always possible. Hence, for our proposed MM scheme, we use a well-known model-free RL algorithm known as Q-learning [6]. Q-learning approximates a value function of each state-action pair through several iterations and learns the optimal policy by using this function.

In our proposed algorithm, we divide the predefined UAV trajectory into discrete states whereby, at each state s , the network can perform an action a and then the UAV moves to next state s' . The reward for taking action a in state s is denoted as $R(s, a)$. The learned action-value $Q(s, a)$ for performing action a in state s is updated using the following rule at each iteration:

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha [R(s, a) + \lambda \max_{a' \in A} Q(s', a')], \quad (5)$$

where $\lambda \in [0, 1]$ is the discount factor, a' is the action that will be taken in the next state s' , and $\alpha \in [0, 1]$ is the learning rate. After the iterative process, the agent will eventually learn the optimal Q-values for each state-action pair, $Q^*(s, a)$ over time. Then, the optimal policy can be obtained by acting greedily in every state by the following equation.

$$\pi^* = \arg \max_{a \in A} Q^*(s, a). \quad (6)$$

We discuss the modeling of state, action, and reward associated with our MM scheme in the following subsection.

A. Simulation Setup and Training

1) *State Representation*: We take the predetermined linear trajectory of the UAV of duration L and divide into δ equal segments where $\delta = \frac{L}{n}$. Here n represents the measurement gap. Note that the given trajectory might already be calculated offline in order to maximize some performance criteria. We denote each segment or waypoint $(x_{\text{uav}}, y_{\text{uav}}, h_{\text{uav}})$ as a unique state. Hence, the number of available states in our setup is δ .

2) *Action Representation*: The network acting as the agent can choose β in the range $[-2, 12]$ with 2° resolution. In other words, the number of available actions is eight. The network can change the values of β by changing the phase of each antenna element remotely during each n .

3) *Reward Model*: To obtain the desired performance, we define a reward function that focuses on maximizing the GUE sum-rate and the RSRPs at the UAV simultaneously along the route. During a flight, the agent needs to tune β of the GBSs while maintaining both of these goals. The weighted combination of RSRP and GUE capacity for taking action a at state s can be expressed as:

$$R(s, a) = w_{\text{rate}} \times C_{s'} + w_{\text{RSRP}} \times \text{RSRP}_{s'}, \quad (7)$$

where $C_{s'}$ and $\text{RSRP}_{s'}$ are the rate of the GUE and the RSRP from the serving GBS at following state s' , respectively. w_{rate} is the weight associated with the GUE rate and w_{RSRP} represents the weight of the RSRP. The weights are chosen such that $w_{\text{rate}} + w_{\text{RSRP}} = 1$.

Algorithm 1 Q-learning for UAV mobility management.

```

1: Input: UAV trajectory, cellular network,  $\epsilon = 1$ ,  $\alpha$ ,  $\gamma$ ,  $Q = \mathbf{0}_{\delta \times 8}$ 
2: repeat (for each iteration):
3:   if  $\epsilon \geq \epsilon_{\min}$ 
4:      $\epsilon = \epsilon \times \eta$  end if
5:   repeat (for each segment of the UAV trajectory):
6:     choose action  $a$  using  $\epsilon$ -greedy policy:
7:     take action  $a$ , calculate  $R(s, a)$  using (7)
8:     update  $Q$ -value using (5) and by calculating the action
       which maximizes  $Q$ -value in the next state
9:      $s \leftarrow s'$ 
10:  until  $s$  is terminal
11: Output:  $Q^*$  table

```

4) *Training Process*: Our offline Q-learning training process consists of multiple iterations or epochs, each with a number of steps δ . At each step, the network tries to obtain the highest reward while also checks for other actions that can improve the estimated future reward. To overcome this exploration-exploitation dilemma, we adopt ϵ -greedy method [33]. The main idea is to choose a random number from $[0, 1]$ and check whether it is smaller than ϵ . If so, the agent takes a random action; otherwise, it goes with the action that has the highest Q -value. We start our first iteration with $\epsilon = 1$ and then reduce ϵ by multiplying a decay factor $\eta = 0.99$ at each iteration. Each iteration starts with $s = 0$ (first waypoint of the UAV trajectory) and finishes when UAV reaches the final location. The pseudo-code of the proposed Q-learning process is presented in Algorithm 1.

It is worth noting that the actions chosen by the agent will impact the RSRP values, which will eventually trigger the HO condition at each state. We count the number of HOs for a given trajectory as follows. At the beginning of the trajectory ($s = 0$), the RSRPs from all the available GBSs is calculated and the UAV associates with the one providing the highest RSRP. We denote the current cell as C_{curr} . For the next waypoint or $s = 1$, C_{curr} remains the same. At $s = 2$, the RSRPs from all GBSs is calculated again and the GBS with the highest RSRP at $s = 2$ is denoted as C_{next} .

If C_{next} is different that C_{curr} and the measured RSRP of C_{next} is higher than that of C_{curr} by HOM, we calculate the RSRP of C_{next} and C_{curr} after travelling TTT s between $s = 1$ and $s = 2$. If the RSRP of C_{next} is higher than that of C_{curr} by HOM, at the moment when TTT ends, HO to C_{next} takes place. Otherwise, the UAV keeps associated with C_{curr} . The same procedure is repeated at each waypoint until UAV reaches the last state.

5) *Analysis of the Algorithm*: In our proposed algorithm, the number of states or the discrete waypoints is finite. The reward function can take values in the range $[0, 1]$ and the set of allowed actions is also finite. Moreover, a decision epoch or iteration finishes when the UAV reaches the final location. Hence, our MM method is an MDP. If we consider $\lambda < 1$ and $\alpha \in [0, 1]$, our proposed Q-learning based MM technique will converge to an optimal action-value function given that Q -values get sufficient number of updates. The complexity of the Q-learning algorithm is $\mathcal{O}(|\mathcal{S}||\mathcal{A}|)$, where $|\cdot|$ denotes the cardinality of a set.

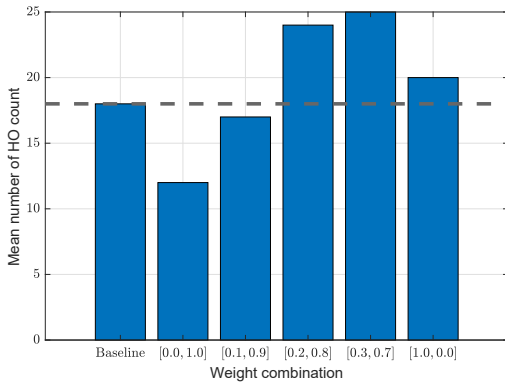


Fig. 3. Average number of HOs per flight for various weight combinations.

TABLE I
SIMULATION PARAMETERS.

| Parameter | Value |
|------------------|--------------|
| P_{gbs} | 46 dBm |
| h_{uav} | 100 m |
| h_{gbs} | 35 m |
| h_{gue} | 1.5 m |
| f_c | 1.5 GHz |
| M | 64 |
| K | 320 |
| v | 120 km/h |
| n | 200 ms [4] |
| TTT, HOM | 160 ms, 3 dB |

V. SIMULATION RESULTS

In this section, we evaluate the performance of our proposed RL-based MM technique. We consider terrestrial networks containing 64 GBSs and 320 GUE in a square area of $4 \times 4 \text{ km}^2$, where the GBSs and the GUE are placed randomly. We consider that the UAV is flying at a speed of 120 km/h in a straight line and the route faces correlated SF. The flying time of the UAV is considered to be 120 seconds. Here, we ignore the take-off and landing time of the UAV as well as the time to reach the altitude h_{uav} . Note that our method can allow more flying time at the expense of higher computational complexity. For Q-learning training purpose, we consider $\alpha = 0.8$, and $\lambda = 0.9$. The Q-learning algorithm is trained offline for 1500 iterations or epochs. We consider 100 random networks for simulation purposes and then compute the average to study the performance of our proposed method. Note that at each realization, the UAV trajectory remains the same; the underlying positions of the GUE and GBSs as well as SF values associated with trajectory change. For each random realization, the RSRP samples associated with each discrete waypoint are linearly transformed to the interval $[0,1]$. We also normalize the GUE rates associated with each action or β for each random network. The simulation parameters and their default values are listed in Table I.

We consider different weight combinations of RSRPs and GUE rates to test the performance of our proposed MM technique and denote weight vector $\mathbf{w} = [w_{\text{rate}}, w_{\text{RSRP}}]$ to represent different weight combinations. For instance, $[1.0, 0.0]$ means that the values of w_{rate} and w_{RSRP} are 1.0 and 0.0, respectively. To compare the performance of our approach, we also consider a baseline in which the value of β is kept fixed at 6° as done for the RMA-AV scenario in [14]. Since

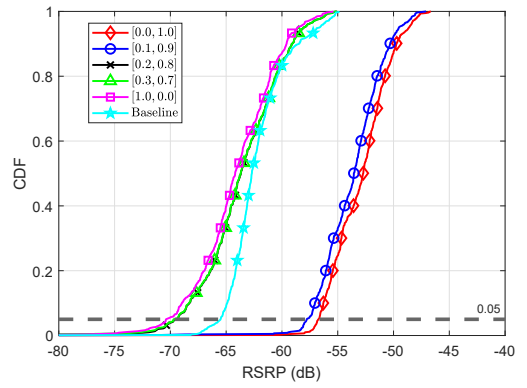


Fig. 4. CDF of RSRP experienced by the UAV for various weight combinations.

Fig. 3 shows the mean number of per-flight HOs completed by the UAV for different weight combinations. We can observe that $[0.1, 0.9]$ and $[0.0, 1.0]$ provide 15% and 40% fewer average HOs, respectively than the baseline scheme where β is kept fixed. On the other hand, a weight vector $[1.0, 0.0]$ scheme yields 11.1% more average HOs compared to the baseline. This is because the more we emphasize the RSRP values, lower values of β are selected which will steer the main beams of the GBSs towards the UAV. This leads to smoother signal coverage in the sky which results in lower HO count. $[0.2, 0.8]$ and $[0.3, 0.7]$ schemes vary β sharply during the flights which create patchy GBS coverage from the UAV's perspective and hence, causes frequent HOs.

TABLE II
DT ANGLE FOR DIFFERENT WEIGHTS.

| Weight vector | mean β |
|---------------|---------------|
| $[0.0, 1.0]$ | -2° |
| $[0.1, 0.9]$ | 6.04° |
| $[0.2, 0.8]$ | 11.97° |
| $[0.3, 0.7]$ | 11.99° |
| $[1.0, 0.0]$ | 12° |

In Table II, we present the per flight mean β selected by the network for different weight combinations. We observe that when more emphasis is placed on the GUE rate, the GBSs will then choose higher values for β . This is consistent with the fact that higher DT angles will decrease the inter-cell interference among the GBSs and will steer the main beam of the antenna towards the GUE with higher gains.

In Fig. 4, we show the cumulative distribution function (CDF) of RSRPs experienced by the UAV in a flight. The weight vector combination $[1.0, 0.0]$ focuses only on the GUE rate and hence, ignores the RSRP performance at the UAV completely. As such, this provides the lowest RSRP for the UAV. On the other hand, $[0.0, 1.0]$ provides the highest RSRP. The difference between the fifth percentile RSRP values associated with these two extremes is around 14 dB. $[0.0, 1.0]$ also provides about 10 dB higher fifth percentile RSRP than the baseline scheme, whereas $[0.1, 0.9]$ provides about 9 dB higher fifth percentile RSRP. The RSRP values associated with the other two weight vectors are very close to each other since they maintain the same average β during the flight duration.

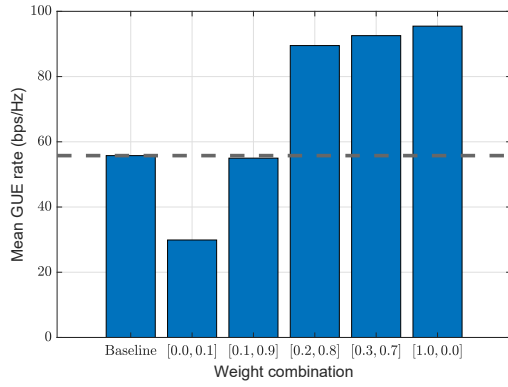


Fig. 5. Average total sum-rate (bps/Hz) of the GUE per flight for various weight combinations.

In Fig. 5, we show the per flight mean GUE rate associated with different weight vectors. As expected, $[1.0, 0.0]$ provides the best GUE rate while its counterpart $[0.0, 1.0]$ provides the lowest one. We can see that $[0.1, 0.9]$ performs very close to the baseline. Overall this weight combination finds a balance between signal quality at the UAV and providing DL coverage to GUE. Moreover, choosing $\mathbf{w} = [0.1, 0.9]$ provides a smaller HO count compared to the baseline scheme.

VI. CONCLUSION

In this paper, we have proposed an RL-based MM framework in a DL cellular network for ensuring better connectivity for both the cellular-connected UAV and the GUE. By exploiting Q-learning, we have provided a flexible technique for finding a balance between these two contradictory goals. The network can trade-off the received signal strength at the UAV with the GUE rate by adjusting the respective weights of these quantities in the reward function and thus by tuning the DT angles accordingly. Our simulation results have demonstrated that the proposed approach can reduce the number of HOs while maintaining good connectivity to the UAV and the GUE, compared to the scenario where the DT angle is always kept fixed. Future extensions can include the case in which multiple UAVs are considered as well as scenarios where the GUE are mobile. In addition, the presence of intelligent reflective surfaces that are facing upward is an interesting future work.

REFERENCES

- [1] G. Geraci, A. Garcia-Rodriguez, L. Galati Giordano, D. Lopez-Perez, and E. Bjornson, "Understanding UAV cellular communications: From existing networks to massive MIMO," *IEEE Access*, vol. 6, pp. 67 853–67 865, 2018.
- [2] R. Amer, W. Saad, and N. Marchetti, "Towards a connected sky: Performance of beamforming with down-tilted antennas for ground and uav user co-existence," *IEEE Commun. Lett.*, pp. 1–1, 2019.
- [3] M. Simsek, M. Bennis, and I. Guvenc, "Mobility management in hetnets: a learning-based perspective," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 26, Feb 2015. [Online]. Available: <https://doi.org/10.1186/s13638-015-0244-2>
- [4] K. Vasudeva, M. Simsek, D. Lopez-Perez, and I. Guvenc, "Analysis of handover failures in heterogeneous networks with fading," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6060–6074, July 2017.
- [5] X. Lin, R. K. Ganti, P. J. Fleming, and J. G. Andrews, "Towards understanding the fundamentals of mobility in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1686–1698, Apr. 2013.
- [6] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

- [7] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145–2146, Nov 1991.
- [8] R. Amer, W. Saad, and N. Marchetti, "Toward a connected sky: Performance of beamforming with down-tilted antennas for ground and uav user co-existence," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1840–1844, Oct 2019.
- [9] R. Amer, W. Saad, and N. Marchetti, "Mobility in the sky: Performance and mobility analysis for cellular-connected UAVs," *IEEE Trans. Commun.*, 2020.
- [10] 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, Apr. 2019.
- [11] R. Amer, W. Saad, B. Galkin, and N. Marchetti, "Performance analysis of mobile cellular-connected drones under practical antenna configurations," in *Proc. IEEE ICC*, Dublin, Ireland, June 2020.
- [12] X. Lin, V. Jainanarayana, S. D. Muruganathan, S. Gao, H. Asplund, H. Maattanen, M. Bergstrom, S. Euler, and Y. E. Wang, "The sky is not the limit: Lte for unmanned aerial vehicles," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 204–210, Apr. 2018.
- [13] X. Lin, R. Wren, S. Euler, A. Sadam, H. Maattanen, S. Muruganathan, S. Gao, Y. E. Wang, J. Kauppi, Z. Zou, and V. Jainanarayana, "Mobile network-connected drones: Field trials, simulations, and design insights," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 115–125, Sept. 2019.
- [14] 3GPP, Technical Specification (TS) 36.777, 2018. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3231>
- [15] H. C. Nguyen, R. Amorim, J. Kovcs, T. B. Srensen, and P. E. Mogensen, "How to ensure reliable connectivity for aerial vehicles over cellular networks," *IEEE Access*, vol. 6, pp. 12 304–12 317, 2018.
- [16] S. Euler, H. Maattanen, X. Lin, Z. Zou, M. Bergstrm, and J. Sedn, "Mobility support for cellular connected unmanned aerial vehicles: Performance and analysis," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.
- [17] Y. Chen, X. Lin, T. Khan, and M. Mozaffari, "Efficient Drone Mobility Support Using Reinforcement Learning," *arXiv e-prints*, p. arXiv:1911.09715, Nov 2019.
- [18] T. Jansen, I. Balan, J. Turk, I. Moerman, and T. Krner, "Handover parameter optimization in lte self-organizing networks," in *Proc. IEEE VTC - Fall*, Sep. 2010, pp. 1–5.
- [19] V. Jainanarayana, H. Ryden, L. Hévizi, A. Jauhari, and M. Cirkic, "5G handover using reinforcement learning," *CoRR*, vol. abs/1904.02572, 2019. [Online]. Available: <http://arxiv.org/abs/1904.02572>
- [20] I. Forkel, A. Kemper, R. Pabst, and R. Hermans, "The effect of electrical and mechanical antenna down-tilting in umts networks," in *Int. Conf. 3G Mobile Commun. Technol.*, May 2002, pp. 86–90.
- [21] O. N. C. Yilmaz, S. Hamalainen, and J. Hamalainen, "Comparison of remote electrical and mechanical antenna downtilt performance for 3gpp lte," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Anchorage, AK, Sep. 2009, pp. 1–5.
- [22] M. Rebato, J. Park, P. Popovski, E. De Carvalho, and M. Zorzi, "Stochastic geometric coverage analysis in mmwave cellular networks with realistic channel and antenna radiation models," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3736–3752, May 2019.
- [23] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2109–2121, Mar. 2018.
- [24] H. Lee and J. Lee, "Joint clock and frequency synchronization for ofdm-based cellular systems," *IEEE Signal Process. Lett.*, vol. 18, no. 12, pp. 757–760, Dec 2011.
- [25] Y. Singh, "Comparison of Okumura, Hata and COST-231 models on the basis of path loss and signal strength," *Int. J. Comp. Appl.*, vol. 59, no. 11, pp. 37–41, Dec. 2012.
- [26] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.
- [27] J. Lee and F. Baccelli, "On the effect of shadowing correlation on wireless network performance," *IEEE INFOCOM*, pp. 1601–1609, 2018.
- [28] T. Klingbrunn and P. Mogensen, "Modelling cross-correlated shadowing in network simulations," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3, Sep 1999, pp. 1407–1411 vol.3.
- [29] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.331, 04 2017, version 14.2.2.
- [30] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.
- [31] I. Siomina, P. Varbrand, and D. Yuan, "Automated optimization of service coverage and base station antenna configuration in umts networks," *IEEE Wireless Communications*, vol. 13, no. 6, pp. 16–25, Dec 2006.

- [32] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, Fourthquarter 2019.
- [33] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Oxford, 1989.