UAV Assisted Cellular Networks With Renewable Energy Charging Infrastructure: A Reinforcement Learning Approach

Michelle Sherman*, Sihua Shao*, Xiang Sun† and Jun Zheng‡
*Department of Electrical Engineering, New Mexico Tech, Socorro, NM 87801.
Email: michelle.sherman@student.nmt.edu, sihua.shao@nmt.edu
†Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131.
Email: sunxiang@unm.edu
‡Department of Computer Science & Engineering, New Mexico Tech, Socorro, NM 87801. Email: jun.zheng@nmt.edu

MILCOM 2021 - 2021 IEEE Military Communications Conference (MILCOM) | 978-1-6654-3956-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/MILCOM52596.2021.9653034

***Abstract*—Deploying unmanned aerial vehicle (UAV) mounted base stations with a renewable energy charging infrastructure in a temporary event (e.g., sporadic hotspots for light reconnaissance mission or disaster-struck areas where regular power-grid is unavailable) provides a responsive and cost-effective solution for cellular networks. Nevertheless, the energy constraint incurred by renewable energy (e.g., solar panel) imposes new challenges on the recharging coordination. The amount of available energy at a charging station (CS) at any given time is variable depending on: the time of day, the location, sunlight availability, size and quality factor of the solar panels used, etc. Uncoordinated UAVs make redundant recharging attempts and result in severe quality of service (QoS) degradation. The system stability and lifetime depend on the coordination between the UAVs and available CSs. In this paper, we develop a reinforcement learning time-step based algorithm for the UAV recharging scheduling and coordination using a Q-Learning approach. The agent is considered a central controller of the UAVs in the system, which uses the ϵ-greedy based action selection. The goal of the algorithm is to maximize the average achieved throughput, reduce the number of recharging occurrences, and increase the life-span of the network. Extensive simulations based on experimentally validated UAV and charging energy models reveal that our approach exceeds the benchmark strategies by 381% in system duration, 47% reduction in the number of recharging occurrences, and achieved 66% of the performance in average throughput compared to a power-grid based infrastructure where there are no energy limitations on the CSs.**

***Index Terms*—UAV, wireless networks, renewable energy, recharging, reinforcement learning.**

I. Introduction

According to the National Oceanic and Atmospheric Administration (NOAA), "the first half of 2020 brought 10 billion-dollar weather disasters, making 2020 the sixth consecutive calendar year where 10 or more billion-dollar weather events occurred, a new record" [1]. These weather-related disasters have led to an inequity of response and affected citizens through preventable loss of life and property. In addition, communication systems can be compromised, leading to longer response times and recovery. UAV-assisted mobile networks with mounted Drone Base Stations (DBSs) can relay network traffic from a local uncompromised Macro Base Station (MBS) to a place of interest (PoI) and provide cellular service to users in that area who may be in need of emergency measures. Utilization of DBS networks will provide rapid information to responders to make informed decisions when entering disaster areas for search and rescue and assessing where there is a priority need for assistance.

Deploying UAV mounted base stations with a renewable energy charging infrastructure in a temporary network (e.g., sporadic hotspots for light reconnaissance mission or disaster-struck areas where regular power-grid is unavailable) is a responsive, flexible, and low-cost solution to provide cellular data access to the user equipment. Figure 1 demonstrates the architecture of the scenario where a free-space optical backhaul link [2] transmits/receives the network traffic to/from the DBS. The RF transceiver onboard the DBS transmits/receives the network traffic to/from the users through the wireless access links. The CSs are powered via solar energy collected by the onboard solar panels and is harvested through a battery storage bank for later use. There are also charging pads on the station used to wirelessly recharge the UAVs when they sit on the platforms.

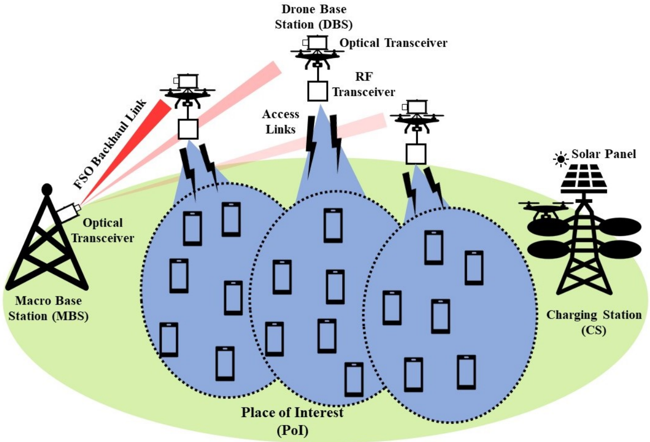

Fig. 1: UAV-assisted mobile network architecture with renewable energy powered recharging infrastructure.

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.

There are three available options for the recharging scheduling of the UAVs: (1) CSs are connected to the power-grid, (2) replenishment stations with UAV replacements or battery replacements [3], and (3) renewable energy CSs. The first option may not be available in the regions under exploration or disaster-struck areas. The second option would require several UAVs or several battery packs to be available on hand, which is a costly solution. The last option, which is also the option we considered in this paper, is the most flexible and cost-effective solution. This option requires the deployment of only a few stations that can charge several UAVs simultaneously.

One of the challenges with solar energy powered CSs is the energy availability limitation. Solar power can be impacted by many environmental factors. This makes the optimization of the recharging coordination between multiple UAVs and multiple CSs more complex. In this paper, we utilize the reinforcement Q-Learning method with a time-step based approach to address the energy availability limitation. We show that the recharging scheduling can be modeled as an ϵ-greedy based method given that the decision policy allows a sufficient exploration-to-exploitation trade-off which, in the long run, will affect the long-term gain. We evaluate the performance of the proposed algorithm using simulations comparing the average throughput received each episode, the number of recharging occurrences per episode, and the time duration of the episode, which is recorded the instant where at least one UAV in the network fully depletes.

In summary, we make three key contributions:

- To the best of our knowledge, the work, in the first time, addresses the scheduling optimization problem of UAV assisted wireless networks with renewable energy CSs as the UAV recharging infrastructure.
- Modeling the energy constraints of the renewable energy CSs in the state space and reward function, we develop a Q-learning based algorithm to optimize the recharging coordination and the average achievable data rate of the users in the PoI.
- Based on experimentally validated energy models from [2], [6], and [13]-[19], extensive simulations are performed to evaluate the impact of the time-step duration on the convergence. Our results confirm that our proposed method out perform the benchmark strategy by 381% and achieves about 66% of recharging option (1) mentioned above where the CSs are connected to the power-grid.

II. Related Work

UAV networks without recharging stations. Trajectory design methods using machine learning approaches are proposed in [4] and [5]. Some works optimize the performance of the network by designing energy-efficient algorithms based on predictions of users' mobility information and geographical fairness. Other work use resource allocation methods to minimize the total energy consumption while satisfying the data rate requirements of the users [6]. [7] proposed a method that enables external input from users to make decisions on their access linkage to enhance the overall achieved throughput. However, none of them take into account the energy limitation on the UAVs.

UAV networks with recharging stations. There have been several works that have looked into utilizing ground [8], [9] and aerial based CSs using RF [10], [11] or optical energy wireless transmission [2]. [11] controls the aerial recharging scheduling of the UAVs in the network via some other "charging host" UAVs. However, the set-up introduces some risks in that there will be a period when the charging host becomes unavailable to the other UAVs in the network and they must stay aloft until the charging host returns. [9] proposes the use of solar-powered CSs to address the challenge of the UAV battery capacity limitation, however, the researchers also assumed the station has a backup option by being connected to the power-grid in case there is not enough power in the battery bank. The work in [8] assumes static CSs with a set battery capacity and proposes a replenishment scheduling policy to minimize the energy consumption by the UAVs while guaranteeing user fairness, however, the simulation duration is very short at only 2 hours.

Other related works. Some other works, such as [4], [8], [11], and [12], develop Q-Learning based methods to solve similar problems using a time-step based approach and are taken as inspiration for the proposed method. Yet, different from these works, our objectives are to increase the life-time of the network by having the algorithm learn how and when to take the best actions at certain time-steps and certain UAV battery thresholds.

III. System Models

A. Scenario

We consider a geographical area of size $X \times Y$ km^2, where a set of K ground users are distributed using a Poisson Point Process. In this network, a set of N UAVs and M static solar powered CSs are deployed uniformly within the geographical area. Each UAV moves between two positions, i.e., its serving spot and its CS assignment given by the Q-learning algorithm. The total service time is given by T and is divided equally into time-steps given by t_S. We assume there is a ground control station (GCS) that is collecting and monitoring the UAV locations, their energy levels, the energy levels of the CSs, as well as controlling the actions of the UAVs.

The locations of the users are given by the horizontal coordinates (x_k, y_k) (m) for $k \in \{1, ..., K\}$. The locations of the UAVs are denoted by (x_n, y_n) (m) for $n \in \{1, ..., N\}$. The locations of the CSs are given by (x_m, y_m) (m) for $m \in \{1, ..., M\}$. The heights of the UAVs and the CSs are denoted as h_{UAV} and h_{CS} (m), respectively. The 3D distance between user k and the UAVs is denoted as $d_{k,n}$ (m). Similarly, the 3D distance between UAV n and CS m is denoted as $d_{n,m}$ (m).

Users are associated with their closest UAVs, which provide them with the highest signal strength. The users are allocated different bandwidth to meet their data rate requirements depending on the channel bandwidth capacity, B (Hz), for each UAV. The current energy of a UAV, per t_S, is given by E_t

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.

(J), where $t \in \{1, ..., (T/t_s)\}$ (sec.). If a UAV is assigned to a CS, the communication links between the UAV and its served users are disconnected until the UAV returns.

B. Channel Model

The air-to-ground (A2G) wireless communication between the UAVs and the users can be divided into line-of-sight (LOS) and non-line-of-sight (NLOS) relations. Probabilistic models have been developed to compute the pathloss obtained for each case [13]. Let $\eta_{k,n}^{LOS}$ be the pathloss in LOS (dB) and $\eta_{k,n}^{NLOS}$ be the pathloss for NLOS (dB) between user k and UAV n. Then we have: $\eta_{k,n}^{LOS} = 20\log_{10}\left(\frac{4\pi f_C d_{k,n}}{c}\right) + \gamma_{LOS}$ and $\eta_{k,n}^{NLOS} = 20\log_{10}\left(\frac{4\pi f_C d_{k,n}}{c}\right) + \gamma_{NLOS}$, where f_C is the carrier frequency (Hz), c is the speed of light (m/s), and γ_{LOS} and γ_{NLOS} account for the mean additional pathloss (dB) found in the LOS and NLOS links, respectively, which are determined by the environment. According to [13] and [14], the access link between user k and UAV n has a LOS probability of $P_{k,n}^{LOS} = \frac{1}{1+\psi e^{-\xi\left(\theta_{k,n}-\psi\right)}}$, where ψ and ξ are the environmental parameters and $\theta_{k,n}$ is the elevation angle (^{o}) between user k and UAV n.

The NLOS probability is $P_{k,n}^{NLOS} = 1 - P_{k,n}^{LOS}$. Therefore, the mean pathloss (dB) between user k and UAV n is given by [13]:

$$\eta_{k,n}^{avg} = \eta_{k,n}^{LOS} P_{k,n}^{LOS} + \eta_{k,n}^{NLOS} P_{k,n}^{NLOS} \text{ (dB)}. \quad (1)$$

C. Data Rate Model

Depending on the available channel bandwidth capacity B, users will be allocated bandwidth to exactly meet their data rate requirement, ϕ_k (Mbps). The achievable data rate, r_k (Mbps), for user k to download data traffic from UAV n is given by the Shannon-Hartley theorem as

$$r_k = b_k \log_2\left(1 + \frac{P_{r,k}}{N_0}\right). \quad (2)$$

b_k is the amount of allocated bandwith to user k (Hz), N_0 is the noise power spectral density measured (Watt/Hz), and $P_{r,k}$ (Watt) is the average received signal power given in [2] as: $P_{r,k} = P_t \times 10^{-\frac{\eta_{k,n}^{avg}}{10}}$, where P_t (Watt) is the transmission power of the access link and $\eta_{k,n}^{avg}$ is the mean pathloss between user k and UAV n given in Eq. (1). Rearranging Eq. (2) and setting $r_k = \phi_k$, the amount of allocated bandwidth is computed by $b_k = \frac{\phi_k}{\log_2\left(1+\frac{P_{r,k}}{N_0}\right)}$.

To achieve the maximum total system capacity, users are associated to their closest UAVs and are allocated bandwidth b_k such that the total amount of bandwidth allocated to the subset of users is no larger that the total amount of available bandwidth B: $\sum_{k\in A_n} b_k \leq B$ where A_n is the set of users associated with UAV n and $A_n \subseteq \{1, ..., K\} \implies \cup_{n=1}^{N} A_n = \{1, ..., K\}$ and $A_n \cap A_{n'} = \emptyset$, where $n \neq n' \; \forall n$ (users are associated with, at most, one UAV).

D. UAV Energy Models

The energy consumption of the UAV depends on the energy needed for: communication, hovering, and mobility mode. Communication energy depends on the transceiver subsystems on the UAV, while hover and mobility energy depend on the motors and weight of the UAV. Considering the proposed system model, hover and mobility modes are considered the dominant energy consumptions of the UAVs in this work.

1) Mobility Energy: The aim is to have the UAVs transition from their hovering spots to their CS assignments and back as quickly as possible as to maximize the system throughput. Thus, we assume all the UAVs are transitioning using a forward flight motion at a constant velocity (V_{UAV}). The thrust (Newtons) for climb, descent, and switch modes are given by: $T_c = W\sin(\theta_{n,m}) + F_{drag}$, $T_d = F_{drag} - W\sin(\theta_{n,m})$, and $T_s = F_{drag}$, respectively, where $W = (M_{UAV} + M_{battery})\,g$, M_{UAV} is the mass of the UAV (kg), $M_{battery}$ is the mass of the UAV battery (kg), g is the gravitational acceleration (kg/m^2), F_{drag} is the drag force (Newtons), and $\theta_{n,m}$ is the climb/descent angle (^{o}) between UAV n and CS m.

From fluid dynamics, the drag force can be expressed as $F_{drag} = \frac{1}{2}\rho V_{UAV}^2 C_D A$, where ρ is the density of air (kg/m^3), V_{UAV} is the constant speed of the UAV (m/s), C_D is the drag coefficient, and A is the cross-sectional area of the UAV (m^2). From conservation of momentum, the required minimum power for propulsion in climb/descent/switch mode is represented by [15]: $P_{min,\{c,d,s\}} = T_{\{c,d,s\}} \times \left(V_{UAV}\sin(\alpha_p) + V_{I,\{c,d,s\}}\right)$ (Watt), where α_p is the pitch or tilt angle (^{o}), $V_{I,\{c,d,s\}}$ is the induced velocity (m/s) in climb/descent/switch mode which can be computed as [15], $V_{I,\{c,d,s\}} = \frac{2T_{\{c,d,s\}}}{\pi R D^2 \rho \sqrt{(V_{UAV}\cos(\alpha_p))^2 + \left(V_{UAV}\sin(\alpha_p) + V_{I,\{c,d,s\}}\right)^2}}$, where R is the number of rotors on the UAV and D is the rotor disc diameter (m).

From the discussed equations, the required power for climb/descent/switch is [15]: $P_{req,mobility,\{c,d,s\}} = P_{min,\{c,d,s\}}/\eta$ (Watt), where η is the overall power efficiency (%). Therefore, the amount of energy consumed by a UAV during transition, per t_S, is: $E_{mobility,\{c,d,s\}} = P_{req,mobility,\{c,d,s\}} t_S$ (J).

2) Hover Energy: It is assumed the UAVs are performing stationary hover modes over their serving spots in no-wind conditions. The power consumption in hover flight mode is adapted from momentum theory by W.J.M. Rankine (1865), A.G. Greenhill (1888), and R.E. Froude (1889) as: $P_{req,hover} = \frac{RT_h^{3/2}}{\sqrt{\frac{1}{2}\pi D^2 \rho}}$ (Watt), where $T_h = W$ is the generated rotor thrust (Newtons). Thus, the amount of energy consumed by a UAV in hover mode per t_S is given by $E_{hover} = P_{req,hover} t_S$ (J).

E. Recharging Station Energy Model

Each CS is powered by means of solar energy via solar panels and battery banks for energy storage. The relationships between a solar panel and the incident solar radiation can be modeled in terms of several angles (^{o}) [16]: ϕ: Latitude - indicates the north/south angular measurement of a point

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.

relative to the Earth's equator; δ: Solar Declination - formed between the equatorial plane and the solar noon; β: Tilt - formed between the solar panel plane and the horizontal; w: Hour angle - expression of time in angular deviation from solar noon; θ: Solar incidence - formed between the incident solar radiation and the normal to the solar panel plane surface.

The solar declination angle is modeled by the following function [16]: $\delta_{\nu} = 23.45 \sin\left(\frac{360(284+\nu)}{365}\right)$, where ν is the day of the year. An equation relating the solar incidence angle (θ_{ν}), the solar panel tilt, the hour angle, and the latitude of the PoI, per t_S, is given by [16]: $\cos\left(\theta_{\nu}(t)\right) = \cos(\phi-\beta)\cos(\delta_{\nu})\cos(w_{\nu}(t)) + \sin(\phi-\beta)\sin(\delta_{\nu})$. $w_{\nu}(t)$ is dependent on the time of the day and the location of the sun. $w_{\nu}(t)$ is divided according to the t_S size and is computed by [16]: $w_{\nu}(t) = 15 \times (Hr(t) - 12)$ for $Hr(t) \leq 12$, or $w_{\nu}(t) = -15 \times (12 - Hr(t))$ for $Hr(t) > 12$, where $Hr(t)$ is the solar time $(Hr(t) \in \{0 : t_S/3600 : 24\})$.

Solar panel performance can be heavily impacted by their orientation and the amount of tilt [17]. [18] presents a method for finding the optimal tilt angle for maximal extraterrestrial radiation over a period of several days, $H_{o,T} = \sum_{\nu=\nu_1}^{\nu_2} H_{o,\nu}$ where $H_{o,\nu}$ is the daily extraterrestrial radiation (Watt/m^2), $T = [\nu_1, \nu_2]$, where the optimal tilt angle is $\beta_{opt,T} = \phi - \tan^{-1}\left[\frac{\sum_{\nu=\nu_1}^{\nu_2} \frac{24}{\pi} G_{sc}\left(1+0.034\cos\left(\frac{2\nu\pi}{365}\right)\right)\sin\left(\delta_{\nu}\right)\frac{w_{s,\nu}\pi}{180}}{\sum_{\nu=\nu_1}^{\nu_2} \frac{24}{\pi} G_{sc}\left(1+0.034\cos\left(\frac{2\nu\pi}{365}\right)\right)\cos\left(\delta_{\nu}\right)\sin\left(w_{s,\nu}\right)}\right]$ for G_{sc} is the solar constant (1367 Watt/m^2) and $w_{s,\nu}$ is the sunset hour angle [16]: $w_{s,\nu} = \cos^{-1}\left(-\tan\left(\phi\right) \times \tan\left(\delta_{\nu}\right)\right)$.

To determine how much energy will be collected at the solar panel surface, [19] proposes an experimentally determined equation to find the amount of sunlight intensity incident on a tilted solar panel, given as:

$$I(t) = G_{sc}\left(0.7^{AM(t)^{0.678}}\right)\sin\left(\alpha_s + \beta_{opt,T}\right) \text{ (Watt/m}^2\text{)}, \tag{3}$$

where $AM(t)$ is the air mass ratio and $\alpha_s = 90^o - \phi + \beta_{opt,T}$ is the solar elevation angle. Eq. (4) is used to compute how much power the CS accumulates for each t_S throughout the period duration T:

$$ch_{rate,CS}(t) = \eta_{sp} \times PR \times A_{sp} \times I(t) \text{ (Watt)}, \tag{4}$$

where η_{sp} is the solar panel efficiency (%), PR is the performance ratio, and A_{sp} is the area of the solar panel (m^2). The performance ratio is a quality factor of the panel that takes into account environmental effects such as degradation.

IV. Recharging Scheduling

A. Q-Learning and ϵ-greedy Method

This paper will use a discrete-time, finite range, Q-Learning algorithm, and ϵ-greedy strategy to make action selections. In the Q-Learning model, the GCS acts as the agent to control the network of N UAVs. At each time-step, the agent observes a state, s_t, from the state space S. The agent can move to the next state s_{t+1} after taking an action, a_t, from the action space A. Upon executing this action, the system receives a reward, r_t, for the state-action pair (s_t, a_t). The reward and state-action pair are used to update a *Q-Table* which is used to determine a decision policy. The Q-values in the Q-Table are updated according to the following formula at each time t [20]:

$$\begin{aligned} Q\left(s_t, a_t\right) \leftarrow\ & Q\left(s_t, a_t\right) + \alpha\big[R(s_t, a_t) + \\ & \gamma \times \max_{\{a \in A\}} Q\left(s_{t+1}, a\right) - Q\left(s_t, a_t\right)\big] \end{aligned} \tag{5}$$

where $\max_{\{a\in A\}} Q\left(s_{t+1}, a\right)$ chooses the action that leads to the maximum Q-value, given the system is in the state s_{t+1}, $\alpha \in [0,1]$ is the learning rate, and $\gamma \in [0,1]$ is the discount factor.

Action selection is determined by the ϵ-greedy method where the agent will try to learn its environment. The process can be outlined as follows: (1) Let $r \in [0,1]$ be a uniform random variable; (2) **if** $r < \epsilon$, **then** select a random action a from action space A, **else** select action a that leads to the maximum Q-Value given the system is in state s: $a \leftarrow \max_{\{a\in A\}} Q\left(s, a\right)$, where $\epsilon \in [0,1]$ is a parameter. Small values of ϵ close to 0 will cause the agent to exploit the information it has learned, while values of ϵ close to 1 will cause the agent to take random actions rather than use acquired past knowledge.

B. Reinforcement Learning

In this paper, the focus of the Q-Learning algorithm is to learn the best coordination between the UAVs in the network and the CSs while taking into account the limited energy available to charge the battery banks at the CSs.

1) State Space: In each time-step, the state $s_t \in S$, is given by the vector

$$\begin{aligned} s_t = \{&B_{UAV1,t}\ \ B_{UAV2,t}\ \ ...\ \ B_{UAVn,t} \\ &B_{CS1,t}\ \ B_{CS2,t}\ \ ...\ \ B_{CSm,t}\}, \end{aligned} \tag{6}$$

where $B_{UAVn,t}$ for $n \in \{1,...,N\}$ represents the battery levels of the UAVs at time-step t (%) and $B_{CSm,t}$ for $m \in \{1,...,M\}$ represents the battery levels of the CSs at time-step t (%). s_t is then a vector with $N+M$ elements. To accelerate the process of learning, the battery levels of the UAVs are divided into four groups. Let $s_{UAVn,t}$ be the state of UAV_n at time t. Then the state of the UAV is given by:

$$s_{UAVn,t} \equiv \begin{cases} 1 & \text{for } B_{UAVn,t} < UAVT_{Low}, \\ 2 & \text{for } UAVT_{Low} \leq B_{UAVn,t} < UAVT_{Mid}, \\ 3 & \text{for } UAVT_{Mid} \leq B_{UAVn,t} < UAVT_{High}, \\ 4 & \text{for } UAVT_{High} \leq B_{UAVn,t}, \end{cases}$$

where $UAVT_{Low}$, $UAVT_{Mid}$, and $UAVT_{High}$ are the lower, middle, and high battery thresholds (%), respectively. Similarly, the battery levels of the CSs are divided into three groups. Let $s_{CSm,t}$ be the state of CS_m at time t. Then the state of the CS is given by:

$$s_{CSm,t} \equiv \begin{cases} 5 & \text{for } B_{CSm,t} < CST_{Low}, \\ 6 & \text{for } CST_{Low} \leq B_{CSm,t} < CST_{High}, \\ 7 & \text{for } CST_{High} \leq B_{CSm,t}, \end{cases}$$

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.

where CST_{Low} and CST_{High} are the low and high battery thresholds (%), respectively. Therefore, there are $4^N \times 3^M$ possible states.

2) Action Space: At each time-step, the UAVs will carry out an action $a_t \in A$, given by

$$a_t = \{A_{UAV1,t} \;\; A_{UAV2,t} \;\; ... \;\; A_{UAVn,t}\} \tag{7}$$

where $A_{UAVn,t} \in \{0, 1, 2, ..., M\}$ for $n \in \{1, ..., N\}$ represents the CS assignment of UAV n. $A_{UAVn,t} = 0$ indicates that UAV n is assigned to its hovering location. Therefore, there are $(M+1)^N$ possible actions.

The charging scheduling is organized on a first-come first-serve basis. Each CS has a lower threshold of $E_{LowT,CS} = (CST_{Low})E_{Cap,CS}$ where $E_{Cap,CS}$ (J) is the total battery bank capacity of the station. This means that once a CS gets below $CST_{Low}\%$ battery capacity, it can only finish satisfying the UAVs that are currently there and will not accept any new assignments. This forces the UAVs to choose between the other ”available” CSs and its hovering spot for future time-steps until the station charges to $\geq CST_{Low}\%$. A CS is considered ”available” if: (1) it can reserve at least some amount of energy to charge a UAV, and (2) there are no more than four UAVs at a station at any given time. If these two conditions are satisfied, the UAV can be charged at that station at a rate of $ch_{rate,UAV}$ (Watt); otherwise, the UAV must select another CS option in the next time-step, or return to hovering.

There are three cases to consider when reserving energy from CSs:

- First, suppose UAV_n is assigned to CS_m at time t. If CS_m has enough energy to fully charge the UAV, then $E_{charge,UAVn} \times (dis_{rate,CSs}/ch_{rate,UAV})$ (J) is reserved from CS_m, where $dis_{rate,CSs}$ is the discharge rate of the CSs (Watt), $E_{charge,UAVn} = E_{Cap,UAV} - (E_{UAVn,t} - E_{mobility,d,UAVn})$ (J), $E_{UAVn,t}$ is the energy of the UAV at t (J), and $E_{Cap,UAV}$ (J) is the battery capacity of the UAVs. The action for UAV_n is held until the charging process is complete. Once the UAV is fully charged, the UAV returns to it's hovering location.
- Now, suppose CS_m does not have enough energy to fully charge the UAV, but it does have enough to charge the UAV another 10% until the station reaches or will go beyond its CST_{Low} threshold. Then, $10\% \times (dis_{rate,CSs}/ch_{rate,UAV})$ is reserved from the station.
- The last case is CS_m cannot provide any energy to charge the UAV. Then, the next action for that particular UAV is limited to only the other available CSs and hovering. This action is selected based on the ϵ-greedy method.
  - The ϵ-greedy decaying function to support convergence of the optimal policy is given by [12]:

    $$\epsilon = \frac{\epsilon_{int}}{(1 + K_{ep}/a)^b}, \tag{8}$$

    where $K_{ep} \in \{1, ..., MaxIter\}$ is the episode number ($MaxIter$ is the maximum number of episodes), ϵ_{int} is the initial ϵ-factor used in episode $K_{ep} = 1$, and a and b are parameters used to alter the steepness and shape of the decaying function.

If multiple UAVs are assigned within the same time-step to the same CS, the order of energy reservations is taken based on which UAV will arrive first (i.e. the transit time). Otherwise, the order is based on the time-step.

As mentioned before, all the UAVs in the network, at each time-step, will either be: (1) hovering and providing communication services to the users in the PoI, (2) transitioning to their assigned CSs, (3) charging, (4) switching to another station, or (5) returning to their hovering spots. An indicator I is used to keep track of what physical state ((1)-(5) above) each UAV is performing at each time-step. Actions that require mobility all take place at the same fixed velocity, V_{UAV}. Using the energy models discussed in Section III-D and the indicator I the number of time-steps the UAV will be in (1)-(5), depending on the action a_t, can be computed.

3) Reward Function: The reward function is computed at the end of each time-step and is given as follows:

$$r_t \equiv \begin{cases} -2\Gamma, & \text{if } \exists \, \{s_{UAVn,t+1} = 1 \mid s_{CSm,t+1} = 5\}, \\ -\Gamma, & \text{if } \exists \, \{A_{UAVn,t} \neq 0\}, \\ \tau_t, & o.w., \end{cases} \tag{9}$$

where Γ is some high value parameter, $\tau_t = \sum_{n=1}^{N} \tau_{n,t}$ is the total throughput received from the N UAVs. Note: $\tau_{n,t} = 0$ if UAV_n is transitioning, charging, returning from the CS, or if the UAV depletes to 0%. $\tau_{n,t} > 0$ only when UAV_n is hovering and $B_{UAVn,t} > 0\%$. Since the achievable data rate for each user is assumed to exactly meet the users' data rate requirements, then the received throughput for each user associated with a UAV is the same as its achievable data rate r_k. The throughput ($\tau_{n,t}$ in Mbps) for UAV_n at time-step t is then $\tau_{n,t} = \sum_{k \in A_n} r_k$. The first penalty in the reward function is defined for cases when the performed action causes $B_{UAV_{n,t}} < UAVT_{Low}$ for any n or $B_{CS_{m,t}} < CST_{Low}$ for any m. The second penalty is defined so the agent learns to minimize the number of recharge occurrences by prioritizing hovering actions. The magnitude of the two penalties ensures that the algorithm finds an efficient tradeoff between the number of recharge occurrences and the lifespan of the network.

If an action causes the battery level of any of the UAVs to reach 0%, then that state-action pair receives a large penalty, and the corresponding Q-value is updated. Once a UAV reaches total depletion at 0% in the updated state s_{t+1}, the time instance t is recorded as $T_{end}(K_{ep})$. The current episode is then terminated and a new episode begins to continue the training process of the Q-Table.

The psuedo-code of the proposed method is summarized in Algorithm 1. The algorithm has three outputs used as metrics to evaluate the system performance in Section V: the average throughput obtained per episode ($AvgT$ in Mbps), the number of recharging occurrences per episode ($RechOccur$), and duration of the episode/system (T_{end} in Days).

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.

```
Algorithm 1: Proposed Q-Learning Algorithm
Input: $N, K, M$, Battery thresholds, $E_{Cap,UAV}$,
  $E_{Cap,CS}$, $ch_{rate,UAV}$, and $ch_{rate,CS}$.
Output: $AvgT$, $RechOccur$, and $T_{end}$.
Initialization: Q-Table, $MaxIter$, $\epsilon(1) = \epsilon_{int}$, and $t_S$.
for $K_{ep} = 1 : MaxIter$ do
    $t = 1$; Set the UAVs and CSs to full charge (100%)
    while $t < T/t_S$ do
        Select an action $a_t \in A$;
        if any UAV begins charging in $t$ then
            Update RechOccur($K_{ep}$);
        end
        Compute the reward $r_t$ using Eq. (9);
        Update UAV and CS energies: $s_{t+1} \in S$;
        Update the Q-Value according to Eq. (5);
        if any($B_{UAVn,t+1} == 0$) then
            Update $Q(s, a)$ with a HIGH penalty;
            break;
        end
        $t = t + 1$;
        Update $ch_{rate,CS}(t)$ using Energy Models;
    end
    Update $\epsilon(K_{ep})$ according to Eq. (8);
    Record $T_{end}(K_{ep})$;
    Compute $AvgT(K_{ep}) = \sum_{t=1}^{T_{end}(K_{ep})} \tau_t$;
end
```

V. Simulation and Performance Evaluation

A. Simulation Setup

In this work[1], we consider $N = 5$ UAVs and $M = 2$ CSs distributed uniformly in a $X \times Y = 3 \times 3$ km^2 grid. The Q-Learning model runs for $MaxIter = 10,000$ episodes. In each episode, the UAVs are assumed to start in their hovering locations with 100% full battery. The battery banks at the CSs also start at 100%. We test the performance of the algorithm for three different time-steps, i.e., $t_S \in \{30, 45, 60\}$ sec. The UAVs provide cellular service in an urban environment with $f_C = 2$ GHz, $B = 20$ MHz, $P_T = 1$ Watt, $N_0 = -104$ dBm/Hz, $\psi = 9.6$, $\xi = 0.28$, $\gamma^{LOS} = 6$ dB, and $\gamma^{NLOS} = 26$ dB. The total time duration, T, is 1 week. The other system model and main simulation parameters are shown in Tables I and II, respectively.

B. Numerical Results and Discussion

For each simulation, the number of recharging occurrences in each episode, and the episode duration results are shown in Figures 2-4, respectively. Figure 5 compares the performance of the proposed algorithm compared to an infrastructure where the CSs are connected to a power-grid, and thus the CSs have no energy limitations.

It can be observed from Figure 2 that the average throughput achieved per episode ($AvgT$) is very low in the first 1500 episodes for all three time-steps. Then, the throughput starts to show some fluctuations between 1500 and 4500 episodes, after which they start to gradually increase and converge. With $\epsilon_{int} = 0.8$, the value of ϵ will decay to 2.544×10^{-5} by the 10000th episode. The converged values are shown by the dashed lines where $RL_{conv} = 137.8$ Mbps for $t_S = 60$ sec., $RL_{conv} = 131.7$ Mbps for $t_S = 45$ sec., and $RL_{conv} = 137.1$ Mbps for $t_S = 30$ sec.

TABLE I: System Model Parameter Values

Notation	Description	Value
ϕ	Latitude of PoI	40^o N
$[\nu_1, \nu_2]$	Days of the Year for Duration Period	$[229, 235]$
PR	Performance Ratio of a Solar Panel	0.8333
A_{sp}	Area of a Solar Panel	1.4424 m^2
η_{sp}	Efficiency of a Solar Panel	35%
M_{UAV}	Mass of a UAV	0.570 kg
$M_{battery}$	Mass of a UAV Battery	0.198 kg
R	Number of Propellers	4
D	Propeller Diameter	0.183 m
A	Projected Area of a UAV	0.046 m^2
C_D	Drag Coefficient of a UAV	1.5
V_{UAV}	Steady-Flight Velocity of a UAV	10 m/s
η	Overall Power Efficiency	0.7
ρ	Air Fluid Density	1.225 kg/m^3
g	Gravity	9.81 kg/m^2
c	Speed of Light	2.997×10^8 m/s
h_{UAV}	Height of a UAV	100 m
h_{CS}	Height of a CS	12 m
$E_{Cap,UAV}$	Energy Capacity of UAV Battery	40.425 Watt-hr
$E_{Cap,CS}$	Energy Capacity of CS Battery	1500 Watt-hr
$ch_{rate,UAV}$	Max. Charging Power of a UAV	38 Watt

TABLE II: Simulation Parameter Values

Notation	Description	Value
T	Time Duration per episode	7 Days
t_S	Time-step per episode	$30, 45, 60$ seconds
$MaxIter$	Number of Episodes	$10,000$
α	Learning Rate	0.9
γ	Discount Factor	0.8
ϵ_{int}	ϵ-greedy initial value	0.8
a, b	ϵ-greedy control factors	$10000, \ 15$

These results are consistent with those shown in Figure 3, which gives the total number of times a UAV began a recharging process during each episode. Initially, due to the high ϵ_{int} value, the UAVs will constantly be recharging, even at high battery levels as the algorithm is exploring actions to take randomly. As the number of episodes increases and the Q-Table is populated with larger quantities, due to the reward received at each time-step, more desirable actions are learned and taken at the appropriate time-step. This causes the number of recharging occurrences to converge and significantly decrease by about 41.8%-47%. Since the UAVs are going to CSs less, then this leads to a higher throughput as they will value hovering and providing cellular service to get a higher reward versus going to a CS too early or too often.

[1] https://github.com/msherman-na24/UAV-ACN-QL.git

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.

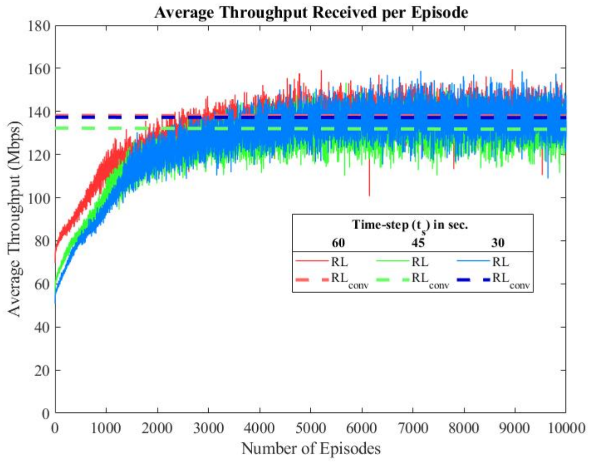

Fig. 2: Evaluation of convergence (shown by the dashed lines) of the average throughput ($AvgT$) for the different time-steps.

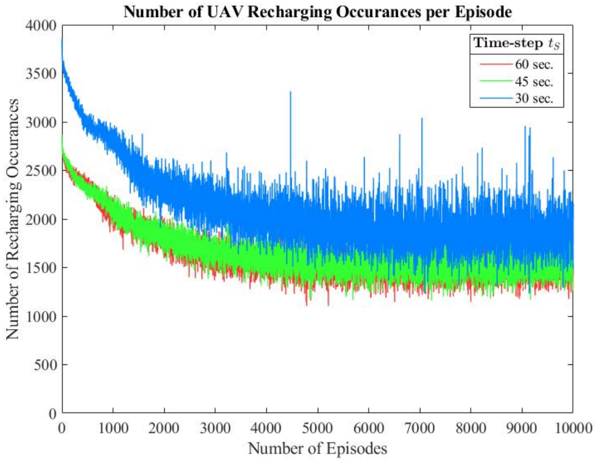

Fig. 3: The number of UAV recharging occurrences ($RechOccur$) as the number of episodes increases for the different time-steps.

Figure 4 shows the time-duration of each episode when at least one UAV fully depleted for each simulation of different time-steps. The dashed line demonstrates the fixed strategy where the UAVs are only sent to their closest CS and there is no coordination between the UAVs and the secondary CS. The fixed strategy constrains the UAVs to only be sent when their battery level is $< 40\%$. Since there is no coordination, even with high efficiency and charging rate of the CSs during the day, one of the CSs is going to drop below the threshold in just over a day (nighttime hours) due to the high energy demand of several UAVs. From the Figure, we can see that the fixed strategy has a time-duration of 1.1 Days, while the proposed algorithm has a time-duration of 4.2 Days for all three time-steps. Over a 381% increase in the life-span of the system compared to the fixed strategy. This implies that the proposed algorithm is utilizing a coordination strategy to schedule the UAVs around the available energy at the CSs. When one station runs below its threshold, UAVs are sent to the other station when they need to be recharged until the first station accumulates enough energy to become available again. We may also notice that there were ten instances where the 30 sec. time-step simulation had a duration of more than 5 Days. Since the 30 sec. time-step simulation had a higher number of time-steps to execute, there were more opportunities for the agent to select random actions and explore compared to the other two simulations, and then exploit this information in the later episodes to improve the performance.

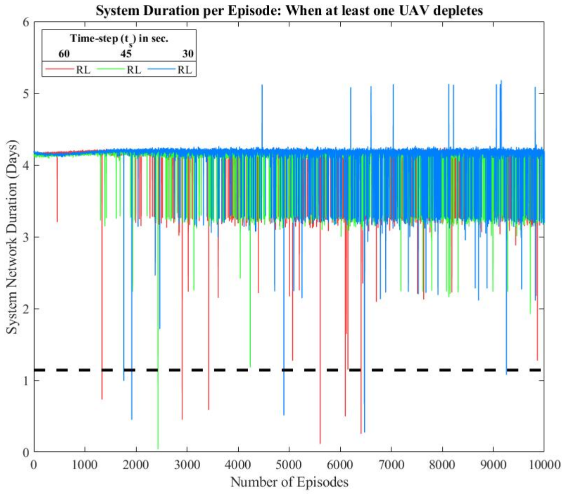

Fig. 4: T_{end} for the proposed algorithm for the different time-steps compared to the fixed strategy (dashed line).

It can also be observed that the proposed algorithm results seemed to reach this average duration in the first few thousand episodes. However, as the number of episodes increases, we see more and more fluctuations in the T_{end} for all three simulations and it does not seem to improve beyond the 4.2 Days. After thorough consideration, we observed that the system is sensitive to the energy threshold levels used by the state space S. Allowing a UAV to go at a lower energy threshold may lead to the system being more prone to depletion instances, even though this will decrease the number of recharging occurrences and thus leading to a higher throughput. This also increases the chance of a UAV encountering a case where there are already four UAVs at a station, but the secondary station is below its threshold so this leaves the UAV struggling to find an available station while getting to dangerously low levels in the meantime. However, if a UAV is constantly being sent to a station at high battery levels, this will have a significant impact on the average throughput. Our goal is to optimize the energy resolution to have better control of the UAV to give it accurate capabilities at different energy levels. The fact that all three time-steps show a high fluctuation of T_{end} as the episode number increases and becomes more dense, this indicates that the agent needs better control to distinguish between different energy levels that are taking the same action strategies determined by the maximum Q-value.

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.

Figure 5 shows a bar plot comparison between the average throughput across all 10000 episodes of the proposed algorithm, for the different time-steps, and an ideal case where there are no limitations on the available energy at a CS at any given time throughput the week. In this case, the UAVs will always be able to charge at their closest CS without having to worry about reserving energy. It is apparent that the different time-steps did not have a significant impact on the simulation results nor on the achieved average throughput for both the proposed algorithm and the power-grid connected infrastructure strategy. However, the proposed algorithm achieved around 66% of the performance of the power-grid strategy.

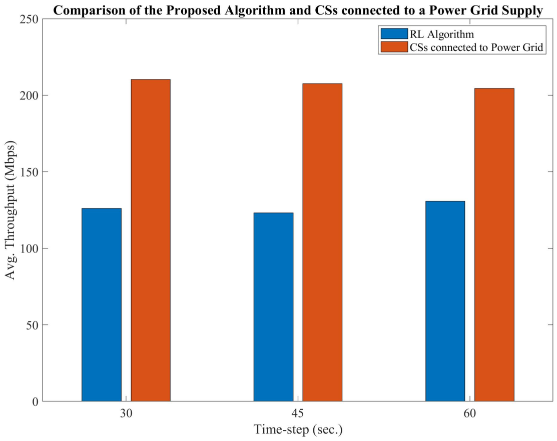

Fig. 5: Average throughput achieved by the proposed algorithm, using each time-step, compared to the case when the CSs are connected to a power-grid.

VI. Conclusion and Future Work

We introduced the concept of using renewable energy powered charging stations and a charging scheduling algorithm to improve the life-span of a temporary UAV-assisted mobile network. We designed a reinforcement Q-Learning algorithm using experimentally derived energy and recharging accumulation models, to closely represent what would occur in practice, with the goal of maximizing the average throughput achieved and minimizing the number of recharging occurrences for a given time duration. Using simulations, it was found that the proposed algorithm can significantly improve the lifespan of benchmark (fixed) strategies. It was also observed that the system is sensitive to the thresholds set for the state space levels of the learning agent, which is heavily impacting the results of the episode time duration. For future work, our objective is to have more precise control of the UAVs in terms of communication and recharging scheduling. This includes diversity of energy levels, trajectory, velocity, channel assignment, etc. Deep reinforcement learning will be investigated to enhance the convergence performance.

Acknowledgment

This work was supported by NSF grant OIA-1757207.

References

[1] "U.S. billion-dollar weather and climate disasters (2020)," https://www.ncdc.noaa.gov/billions/DOI:10.25921/stkw-7w73, accessed: 2021-03-30.

[2] D. Wu, X. Sun, and N. Ansari, "An FSO-based drone assisted mobile access network for emergency communications," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 1597–1606, 2019.

[3] B. Galkin, J. Kibilda, and L. A. DaSilva, "Uavs as mobile infrastructure: Addressing battery lifetime," *IEEE Communications Magazine*, vol. 57, no. 6, pp. 132–137, 2019.

[4] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7957–7969, 2019.

[5] L. Wang, K. Wang, C. Pan, W. Xu, N. Aslam, and L. Hanzo, "Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, pp. 73–84, 2021.

[6] W. Ding, Z. Yang, M. Chen, J. Hou, and M. Shikh-Bahaei, "Resource allocation for UAV assisted wireless networks with QoS constraints," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–7.

[7] Y. Cao, L. Zhang, and Y.-C. Liang, "Deep reinforcement learning for multi-user access control in UAV networks," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[8] H. Qi, Z. Hu, H. Huang, X. Wen, and Z. Lu, "Energy efficient 3-D UAV control for persistent communication service and fairness: A deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 53 172–53 184, 2020.

[9] P.-V. Mekikis and A. Antonopoulos, "Breaking the boundaries of aerial networks with charging stations," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[10] S. A. Hoseini, J. Hassan, A. Bokani, and S. S. Kanhere, "Trajectory optimization of flying energy sources using Q-Learning to recharge hotspot uavs," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2020, pp. 683–688.

[11] J. Xu, K. Zhu, and R. Wang, "RF aerially charging scheduling for UAV fleet : A Q-Learning approach," in *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, 2019, pp. 194–199.

[12] S. Shao, G. Liu, A. Khreishah, M. Ayyash, H. Elgala, T. D. Little, and M. Rahaim, "Optimizing handover parameters by Q-Learning for heterogeneous radio-optical networks," *IEEE Photonics Journal, PP (99)*, pp. 1–1, 2019.

[13] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal lap altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[14] Y. Qin, M. A. Kishk, and M.-S. Alouini, "Performance evaluation of UAV-enabled cellular networks with battery-limited drones," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2664–2668, 2020.

[15] J. Stolaroff, C. Samaras, E. R. O'Neill, A. Lubers, A. S. Mitchell, and D. Ceperley, "Energy use and life cycle greenhouse gas emissions of drones for commercial package delivery," *Nature Communications*, vol. 9, 2017.

[16] J. Duffie and W. Beckman, *Solar Engineering of Thermal Processes*. John Wiley and Sons, Ltd, 2013, ch. 1, pp. 3–42. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118671603.ch1

[17] M. Mamun, M. Islam, M. Hasanuzzaman, and J. Selvaraj, "Effect of tilt angle on the performance and electrical parameters of a PV module: Comparative indoor and outdoor experimental investigation," *Energy and Built Environment*, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666123321000179

[18] R. Abdallah, A. Juaidi, S. Abdel-fattah, F. Manzano-Agugliaro, and R. Khaldi, "Estimating the optimum tilt angles for south-facing surfaces in palestine," *Energies*, vol. 13, 02 2020.

[19] R. B. Nazmul, "Calculating optimum angle for solar panels of dhaka, bangladesh for capturing maximum irradiation," in *2017 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2017, pp. 25–28.

[20] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. The MIT Press, 2018.

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on February 09,2022 at 19:18:41 UTC from IEEE Xplore. Restrictions apply.