

Optimal Charging Profile Design for Solar-Powered Sustainable UAV Communication Networks

Longxin Wang*, Saugat Tripathi[†], Ran Zhang[†], Nan Cheng*, and Miao Wang[†]

* School of Telecommunication Engineering, Xidian University, Xi'an, China

[†] Department of Electrical and Computer Engineering, Miami University, Oxford, USA

Email: *wanglx19@stu.xidian.edu.cn, *dr.nan.cheng@ieee.org, [†]{tripats, zhangr43, wangm64}@miamioh.edu

Abstract—This work studies optimal solar charging for solar-powered self-sustainable UAV communication networks, considering the day-scale time-variability of solar radiation and user service demand. The objective is to optimally trade off between the user coverage performance and the net energy loss of the network by proactively assigning UAVs to serve, charge, or land. Specifically, the studied problem is first formulated into a time-coupled mixed-integer non-convex optimization problem, and further decoupled into two sub-problems for tractability. To solve the challenge caused by time-coupling, deep reinforcement learning (DRL) algorithms are respectively designed for the two sub-problems. Particularly, a relaxation mechanism is put forward to overcome the "dimension curse" incurred by the large discrete action space in the second sub-problem. At last, simulation results demonstrate the efficacy of our designed DRL algorithms in trading off the communication performance against the net energy loss, and the impact of different parameters on the tradeoff performance.

I. INTRODUCTION

In future mobile communications networks, UAVs equipped with wireless transceivers can be exploited as mobile base stations to provide highly on-demand services to the ground users, forming UAV-based communication networks (UCNs). With the advantages of flexible 3D mobility, higher chance of Line-of-Sight communication channels, and lower deployment and operational cost, UCNs have received substantial research attention from various aspects [1], [2]. However, the existing works mostly focus on UAV control considering a fixed set of UAVs. Few works have investigated how the network should optimally respond when the UAV crew dynamically change. On one hand, UAVs are powered by batteries. Some UAVs will run out of battery during the service and have to quit the network for charging. On the other hand, supplemental UAVs can be dispatched to easily join the existing crew to enhance the network performance. Therefore, it is indispensable to design novel responsive regulation strategies capable of optimally handling a UAV crew that may change dynamically.

To this end, we proposed in [3], [4] a responsive UAV trajectory control strategy to maximize the accumulated number of served users over a time horizon where at least one UAV quits or joins the network. Nevertheless, no matter how good such responsive strategies can be, they are by nature passive reaction strategies which can only accept and passively react to the change rather than proactively control the change. Solar charging makes the proactive control possible. The chance

lies in that the user traffic demand in an area is usually time-varying. When the demand is low, some UAVs can be deliberately dispatched to elevate high to get solar charged even if they are not in bad need of charging. They can be called back later to replace other UAVs or meet the increased user demand. In this manner, the network is able to take charge of the change in the serving UAV crew, and a solar-powered sustainable (SPS) UAV network can be established.

UAV communications with solar charging have been studied by some pioneering works. With a single solar-powered UAV, [5] developed an optimal 3D trajectory control and resource allocation strategy, [6] studied the problems of energy outage at UAV and service outage at users by modeling solar and wind energy harvesting, and [7] proposed a novel power cognition scheme to intelligently adjust the energy harvesting, information transmission, and trajectory to improve UAV communication performance. For multi-UAV networks, the work [8] studied joint dynamic UAV altitude control and multi-cell wireless channel access management to optimally balance between solar charging and communication throughput improvement. The work [9] analytically characterized the user coverage performance of a UAV network based on a harvest power model and 3D antenna radiation patterns. Although solar charging is exquisitely integrated to fuel the UAV communications, most of the related works do not take into account the time-variability of solar radiation or user traffic demand, which, however, is usually the case in practice. To achieve day-scale sustainability, it is essential to consider these time variation when designing the UAV control strategies.

Therefore, in this work, we investigate the optimal solar charging strategy design for a UCN, considering time-varying solar radiation and user data traffic demand. The strategy aims to optimally trade off over a time horizon between maximizing the accumulated user coverage and minimizing the net energy loss, subject to the constraints of UAV sustainability and user service requirements. The net energy loss is defined as the difference between the total energy harvest and the total energy consumption. As far as we know, this work is the first to jointly consider time-varying solar radiation and user service demand at a day-scale in a solar-powered self-sustainable UAV network. Specifically, our contributions are three-fold.

- The studied problem is first formulated into a time-coupled mixed-integer non-convex optimization problem.

To make the problem tractable, the original problem is decoupled into two sub-problems, one obtaining the mapping between the number of serving UAVs and the number of served users in each time slot, and the other handling the time-variability of solar radiation and user service demand.

- To tackle the challenge caused by time coupling, deep reinforcement learning (DRL) algorithms are developed to solve the two sub-problems. Particularly, a relaxation mechanism is designed to relieve the “dimension curse” caused by the large discrete action space in the second sub-problem.
- Simulations are conducted to demonstrate the efficacy of the proposed learning algorithms and the impact of different parameters on the tradeoff performance.

remainder of the paper is organized as follows. Section II depicts the system model. Section III presents the problem formulation and decomposition. Section IV details the proposed DRL algorithms and its relaxation. Section V provides the numerical results. Section VI concludes the paper.

II. SYSTEM MODEL

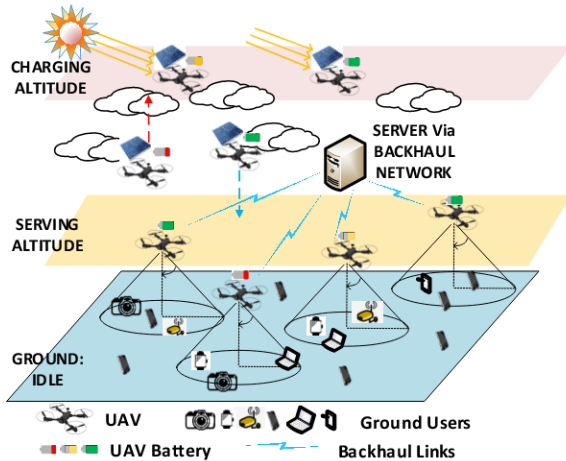


Fig. 1: System model of SPS UAV communication network.

A. Network Model

As shown in Fig. 1, we consider N solar chargeable UAVs, denoted as $\mathcal{S}_{UAV}$, providing communication services to (i.e., serve) a target area. All UAVs can communicate with a server via backhaul networks (e.g., a satellite or cellular network). Each UAV concentrates its transmission energy within an aperture underneath it, forming a ground coverage disk. UAVs are mostly at three altitudes: ground, the serving altitude (H_{Srv}) and the charging altitude (H_{Chg}). When on the ground, UAVs consume negligible power only for messaging with the server. UAVs only serve and get charged at the fixed altitude H_{Srv} and H_{Chg} , respectively. H_{Srv} is low to maintain good

UAV-user communication quality, while H_{Chg} is right above the upper boundary of the clouds to minimize the attenuation of solar radiation due to clouds. The consideration of only charging UAVs at H_{Chg} is justified as follows. According to [10], the solar radiation is attenuated exponentially with the thickness of clouds between the sun and solar panel, leading to only $\sim \frac{1}{10}$ after the first 300 meters. When it does not take long (e.g., one minute or two) for a UAV to move vertically through 300 meters, UAVs can be reasonably set for charging at a fixed altitude just above the clouds.

The time horizon T is equally divided into time slots indexed by t . In any time slot, a percentage p of the users are randomly distributed in proximity to some hotspot centers while the remaining are uniformly distributed throughout the area. The numbers and spatial distributions of the users and hotspots are deemed unchanged within a time slot but may vary with t . The dynamics of the user distribution are known to the server to obtain an offline UAV charging strategy. The well-trained strategy will be executed relying on the server-UAV communications via the backhaul links.

B. Spectrum Access

Users access the UAV spectrum following LTE Orthogonal Frequency Division Multiple Access (OFDMA) [11], which assigns different users of one UAV at least one orthogonal Resource blocks (RBs) such that they do not interfere with each other. A heuristic 2-stage user association policy is adopted. In each time slot, users send connection requests in stage I to the serving UAVs that provides the best SINR (can be measured via reference signaling), and each UAV admits the users with the best SINR values based on its bandwidth. In stage II, the rejected users then chooses the UAV with the next best SINR and is admitted if the UAV has available bandwidth. The stage II procedure is repeated for each user without association until it is admitted or has no available UAV to send requests to. Each user has a minimum throughput requirement r_u . When a UAV i admits a user u , the number of RBs assigned to the user, n_{iu}^{RB} , should satisfy

$$n_{iu}^{RB} W^{RB} \log_2 \left(1 + \frac{P_t G_{iu}}{n_0 + \sum_{j \in \mathcal{S}_{u'} \setminus \{i\}} P_t G_{ju}} \right) \geq r_u \quad (1)$$

where $G_{iu} = 10^{-PL_{iu}/20}$,
 $PL_{iu} = 20 \log_{10} \left(\frac{4\pi f_c d_{iu}}{c} \right) + \eta$ (dB).

In Eq. (1), W^{RB} is the bandwidth per RB, P_t is the transmit power spectrum density (psd) of UAVs, n_0 is the noise psd, $\mathcal{S}_{u'}$ denotes the set of UAVs that can cover user u , and G_{iu} is the UAV-to-user channel gain as a function of the center frequency f_c , distance d_{iu} between UAV i and user u , light speed c and a line of sight (LoS) related parameter η [12].

C. Energy Model

We follow the work in [5] to model the kinematics power consumption for the UAVs. For a UAV flying at a level speed v_{lv} and a vertical speed v_{vt} , the kinematics power consumption is modeled as Eq. (2). In Eq. (2), P_{lv} , P_{vt} and P_{drag} denotes the power consumption of level flight, vertical flight and blade

drag profile power, respectively, W is the UAV weight, ρ is the air density, A is the total area of the UAV rotor disks, C_{D0} is the profile drag coefficient, σA is the total blade area, and v_T is the blade tip speed. Note that v_{vt} is positive for UAV climbing, and negative for UAV landing.

$$\begin{aligned} P_{kine} &= P_{lv} + P_{vt} + P_{drag}, \\ \text{where } P_{lv} &= \frac{W^2}{\sqrt{2\rho A} \sqrt{v_{lv}^2 + \sqrt{v_{lv}^4 + 4V_h^4}}}, \\ P_{vt} &= Wv_{vt}, \\ P_{drag} &= \frac{1}{8}C_{D0}\rho \cdot \sigma A ||v_T||^3 \\ V_h &= \sqrt{\frac{W}{2\rho A}}. \end{aligned} \quad (2)$$

In addition to the kinematics power consumption, UAVs spend power on communication and on-board operations like computing, which are denoted as P_{tx} and P_{static} , respectively. Thus the total power consumption of a UAV is given as

$$P_{Tot} = P_{kine} + P_{tx} + P_{static}. \quad (3)$$

Note that P_{kine} is usually several hundred watts. The transmission power of a small BS covering hundreds of meters typically falls between 0.25W and 6W [13]. The operational power consumption is also in single-digit watts. Thus, P_{tx} and P_{static} are usually neglected in practice.

The solar radiation intensity above the clouds varies with time in a day. We follow the model in [6] to characterize the average intensity as

$$I_{rad}(t) = \max\{0, I_{max}(-1/36t^2 + 2/3t - 3)\}, 0 \leq t < 24, \quad (4)$$

where t represents hour t , and I_{max} denotes the maximum intensity during a day. The harvested solar power is then calculated as

$$P_h(t) = \begin{cases} A_c \frac{\eta_c}{K_c} I_{rad}(t)^2, & 0 < I_{rad}(t) < K_c, \\ A_c \eta_c I_{rad}(t) & I_{rad}(t) \geq K_c, \end{cases} \quad (5)$$

where A_c is the solar panel area, η_c is the charging efficiency coefficient, and K_c is an intensity threshold.

III. PROBLEM FORMULATION AND DECOMPOSITION

The objective is to achieve the optimal tradeoff among maximizing the total number of served users over the time horizon T , maximizing the total harvested solar energy, and minimizing the total energy consumption of the UAV network. The optimization is subject to the network sustainability constraints and user traffic demand requirements. With the above considerations, the problem formulation is given as P_1 .

In Problem P_1 , the decision variables include whether a UAV should land, go serving or go charging at any time slot t , i.e., $\hat{\mathbf{a}}_t = (a_{1,t}, \dots, a_{N,t})$, and the horizontal positions of the UAVs that go serving in any time slot t , i.e., $\hat{\mathbf{p}}_t = (a_{k_1(t),t}, \dots, a_{k_M(t),t})$, where $k_m(t), m \in \{1, \dots, M\}$ index the serving UAVs at time slot t . Part A_1 denotes the amount of energy harvest from solar charging for UAV i at time slot t . This part is determined by $a_{i,t}$ and $a_{i,t-1}$ since a UAV takes some time to move from the last altitude to the current one, the harvest solar power $P_h(t)$, and the battery

residue at the end of time slot $t-1$, i.e., E_{t-1}^{res} , since the battery capacity may be reached during charging. Part A_2 represents the energy consumption of UAV i at t , which is also determined by $a_{i,t}$, $a_{i,t-1}$, and E_{t-1}^{res} . In Part A_3 , $\hat{\mathcal{S}}_t^u(\cdot)$ is the set of users that are admitted and served by all the UAVs at t , which is a function of $\hat{\mathbf{a}}_t$ and $\hat{\mathbf{p}}_t$. The constant C is the coefficient balancing the weights between the user coverage and the energy gain and losses.

$$\begin{aligned} \max_{\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t} \quad & \sum_{t=1}^T \left\{ C \sum_{i \in \mathcal{S}_{UAV}} \underbrace{[E_h(a_{i,t}, a_{i,t-1}, P_h(t), E_{i,t-1}^{res})]}_{A_1} \right. \\ & \left. - \underbrace{E_c(a_{i,t}, a_{i,t-1}, E_{i,t-1}^{res})}_{A_2} + \underbrace{|\hat{\mathcal{S}}_t^u(\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t)|}_{A_3} \right\} \\ \text{s.t.} \quad & E_{i,t}^{res} \geq E_{min}(a_{i,t}), \forall i \text{ and } \forall t; \\ & |\hat{\mathcal{S}}_t^u(\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t)| \geq p_{min} |\mathcal{S}_t^u|, \forall t; \\ & \text{Eq.(1), } \forall u \in \hat{\mathcal{S}}_t^u(\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t) \text{ and } \forall t. \end{aligned} \quad \begin{aligned} & (P_1) \\ & (C1.1) \\ & (C1.2) \\ & (C1.3) \end{aligned}$$

Constraint C1.1 represents the network sustainability requirements. The battery residue of any UAV i at any t should be no smaller than an altitude-dependent threshold $E_{min}(a_{i,t})$. This is to make sure at the end of each t , each UAV has enough energy to elevate to H_{Chg} for charging in future slots to avoid completely leaving the crew. Thus,

$$E_{min}(a_{i,t}) = \Delta H / v_{up} \cdot (P_{Tot}|_{v_{lv}=0, v_{vt}=v_{up}}), \quad (6)$$

where ΔH takes H_{Chg} when $a_{i,t}=0$, $H_{Chg}-H_{Srv}$ when $a_{i,t}=1$, and 0 when $a_{i,t}=2$. Constraints C1.2 and C1.3 represent the user data traffic demand requirements. C1.2 requires that the percentage of served users at any t should be no less than p_{min} given the total number of users $|\mathcal{S}_t^u|$. C1.3 requires that the individual user traffic demand r_u should be satisfied for any served users at any t .

Problem P_1 is a mixed integer nonlinear non-convex optimization problem with nonlinear constraints. The items of different t in the objective function are temporally coupled through UAV battery residue. These facts make the sequential decision problem intractable. To this end, we decouple P_1 into two sub-problems P_2 and P_3 , each of which can be solved by means of DRL algorithms.

$$\max_{\hat{\mathbf{p}}_t} \quad |\hat{\mathcal{S}}_t^u(\hat{\mathbf{p}}_t, N_{UAV}^{Srv})|, \quad \forall t \quad (P_2)$$

$$\text{Eq.(1), } \forall u \in \hat{\mathcal{S}}_t^u(\hat{\mathbf{p}}_t, N_{UAV}^{Srv}). \quad (C2.1)$$

In the first sub-problem P_2 , given the user distribution at each t and the number of UAVs in service N_{UAV}^{Srv} , the total number of users that can be served is maximized via determining the optimal positions $\hat{\mathbf{p}}_t^*(N_{UAV}^{Srv})$ as a function of N_{UAV}^{Srv} and the user distribution.

In the second sub-problem P_3 , the achieved mapping from P_2 between the maximum number of served users $|\hat{\mathcal{S}}_t^u(\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t^*(N_{UAV}^{Srv}))|$ and N_{UAV}^{Srv} is exploited. The same objective as in P_1 is maximized via only optimizing $\hat{\mathbf{a}}_t$. The

relationship between N_{UAV}^{Srv} and $\hat{\mathbf{a}}_t$ is given in Eq. (7), where $I(\cdot)$ is a binary indicator taking 1 if the inside condition is true and 0 otherwise.

$$\max_{\hat{\mathbf{a}}_t} \sum_{t=1}^T \left\{ C \sum_{i \in \mathcal{S}_{UAV}} [E_h(a_{i,t}, a_{i,t-1}, P_h(t), E_{i,t-1}^{res})] - E_c(a_{i,t}, a_{i,t-1}, E_{i,t-1}^{res}) + |\hat{\mathcal{S}}_t^u(\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t^*(N_{UAV}^{Srv}))| \right\} \quad (P_3)$$

$$\text{s.t. } E_{i,t}^{res} \geq E_{min}(a_{i,t}), \forall i \text{ and } \forall t; \quad (C3.1)$$

$$|\hat{\mathcal{S}}_t^u(\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t^*(N_{UAV}^{Srv}))| \geq p_{min} |\mathcal{S}_t^u|, \forall t; \quad (C3.2)$$

$$N_{UAV}^{Srv} = \sum_{i \in \mathcal{S}_{UAV}} I(a_{i,t} == 1), \forall t. \quad (7)$$

IV. DESIGN OF DEEP REINFORCEMENT LEARNING ALGORITHMS

In this section, design of the DRL algorithms for solving P_2 and P_3 are elaborated. For P_2 , we reuse the algorithm designed in our previous work [4] to achieve the mapping between the number of UAVs in service (N_{Srv}^{UAV}) and the total number of served users ($|\hat{\mathcal{S}}_t^u|$) given a certain user distribution in hour t . Based on the time-dependent mappings, the DRL algorithm design for solving P_3 is emphatically presented.

A. DRL for Solving P_2

Our previous work [4] considered a set of UAVs flying at a fixed height providing communication services to the ground users with minimum throughput requirement. We considered dynamic UAV crew change during the training due to battery depletion or supplementary UAV join-in. A DDPG algorithm was designed to maximize the user satisfaction score via obtaining the optimal UAV trajectories during the steady period without crew change and the transition period when crew changes. With the following simplifications, the proposed algorithm can be fit to P_2 . Crew change is not considered so that the state space can be cut down to only including the UAV positions. The action space remains unchanged allowing a UAV to go any direction with a maximum distance d_{max} per step. The reward function changes from step-wise user satisfaction score to step-wise total number of served users. The closest-SINR based user association policy is replaced with that elaborated in Subsection II-B.

B. DRL for Solving P_3

P_3 exploits the mappings between $|\hat{\mathcal{S}}_t^u|$ and N_{Srv}^{UAV} in different hours obtained via P_2 , and aims to maximize its objective function via optimizing the UAVs' charging profiles in the considered time horizon. In each hour t , the DRL agent needs to determine whether a UAV should go charging, go serving or go to the ground for energy saving, i.e., $\hat{\mathbf{a}}_t$, based on the UAVs' current battery residues and altitudes, solar radiation intensity, and user traffic demands. When designing the DRL algorithm, we consider the varying solar radiation and user traffic demands as the dynamics of the underlying environment. The key components of the algorithm are designed as follows.

1) *State Space*: Battery residue of a UAV is a critical factor in determining its next move, thus $\{E_{i,t}^{res}\}, \forall i \in \mathcal{S}_{UAV}$ are included into the state space, denoting the residue battery of UAV i at the beginning of hour t . The current UAV altitude is another in-negligible factor as the altitude changing will incur wear energy consumption which may accumulate to significantly impact the overall scheduling. Minimizing the unnecessary altitude changes for UAVs while satisfying the constraints will contribute positively to the optimization objective. Therefore, $\{H_{i,t}\}, \forall i \in \mathcal{S}_{UAV}$ are embraced in the state space, which takes values H_{Chg} , H_{Srv} , or 0. Lastly, the hour indexing t needs to be considered to capture the dynamics of the environment (e.g., solar radiation and user traffic demand) so that different actions may be taken at different t even if the rest of the states are same. The complete state space is given below, with a cardinality of $2N + 1$.

$$\mathbf{S}_t = \{E_{i,t}^{res}, H_{i,t}, t\}, \quad \forall i \in \mathcal{S}_{UAV}. \quad (8)$$

2) *Action Space*: The decision variables of subproblem P_3 is $\hat{\mathbf{a}}_t = (a_{1,t}, \dots, a_{N,t})$, denoting the altitudes that each UAV will go to at the beginning of the current hour t . The action space is directly defined as $\mathbf{A}_t = \{a_{i,t}\}, \forall i \in \mathcal{S}_{UAV}$, which takes value 0 if the UAV goes to the ground, 1 if the UAV goes serving, and 2 if the UAV goes charging. The cardinality of the action space is 3^N .

3) *Reward Function Design*: The reward function r_t consists of three parts. The first part $r_{1,t}$ corresponds to the constraints of P_3 . When any UAV breaks the sustainability constraint (C3.1), a constant penalty $p_{C1} < 0$ is applied. When constraint C3.2 is broken, i.e., the total number of serving UAVs N_{UAV}^{Srv} cannot result in a minimum user service rate p_{min} , a constant penalty $p_{C2} \in (p_{C1}, 0)$ is applied. In addition, when N_{UAV}^{Srv} is larger than the minimum number of serving UAVs that result in 100% user service rate, a reward of 0 is applied to prevent service over-provisioning.

The second part $r_{2,t}$ corresponds to the maximization of the total number of served users over the entire time horizon. Thus, $r_{2,t}$ is set to be directly equal to $|\hat{\mathcal{S}}_t^u(\hat{\mathbf{a}}_t, \hat{\mathbf{p}}_t^*(N_{UAV}^{Srv}))|$. The third part $r_{3,t}$ corresponds to the maximization of the difference between the total harvested energy and the total consumed energy. Due to the time-varying solar radiation intensity, it is beneficial for a UAV to land to the ground if it does not serve during some hours of a day (e.g., in the night or around sunset/sunrise), while it is beneficial to go charging during other hours of a day. In the former case, positive reward is given for UAVs going to the ground to promote energy saving, whereas in the latter case, positive reward is given for UAVs going charging to encourage energy harvesting. Therefore, r_3 is designed as

$$r_{3,t} = \begin{cases} c_1 \cdot N_{UAV}^{Gnd}, & \text{if landing is beneficial at } t; \\ c_2 \cdot N_{UAV}^{Chg}, & \text{if charging is beneficial at } t, \end{cases} \quad (9)$$

where c_1 and c_2 are reward coefficients to tradeoff between A_1 and $A_3 - A_2$ in P_1 , replacing coefficient C for A_1 . The total instantaneous reward r_t is $r_t = r_{1,t} + r_{2,t} + r_{3,t}$.

4) *Relaxation of the Discrete Action Space*: As the state space $\mathbf{S}_t$ is continuous and discrete mixed, and the action space $\mathbf{A}_t$ is discrete, Deep Q learning (DQL) algorithm is typically exploited. However, the cardinality of the action space, i.e., 3^N , increases exponentially with the total number of UAVs N . For an instance of $N=15$, the total number of possible aggregate actions over all UAVs will be $3^{15} \approx 1.4e^7$. As the number of outputs of the Deep Q network (DQN) is equal to the total number of possible actions, the resultant DQN will be prohibitively complicated, not to mention the number of hours in the considered time horizon. Therefore, DQL is technically feasible, but practically impossible.

Inspired by [14], we relax the original discrete action space into a continuous space and obtain the UAV charging profile $\hat{\mathbf{a}}_t, \forall t \in T$ by means of DDPG. Each action $a_{i,t}$ is relaxed from discrete values $\{0, 1, 2\}$ to a continuous range $(-0.5, 2.5)$. Hence, the relaxed action space becomes $\hat{\mathbf{A}}_t = \{a_{i,t}\} \in (-1.5, 2.5)^N$. With DDPG, the number of outputs of the actor network is equal to the dimension of the action space, i.e., N , which only increases linearly with N rather than exponentially as 3^N in DQN. Each time when an aggregate action is determined by the actor network and added with noise, the action will then be discretized to the closest value in $\{0, 1, 2\}$. The discretized action will be the actual action applied to the current state and stored in the experience replay buffer. In this manner, the complexity of the exploration can be significantly reduced.

V. NUMERICAL RESULTS

A. Simulation Setup

For subproblem P_2 , we reuse the simulation setup and parameter configurations in our previous work [4] to obtain the hourly mapping between the number of UAVs and the maximal number of served users. For subproblem P_3 , the environment parameters and the RL parameters are summarized in Table I and II, respectively. A 24-hour time horizon is considered per episode with each hour being a step. The training is conducted using Reinforcement Learning Toolbox of Matlab 2022a on a Windows 10 server with Intel Core i9-10920X CPU @ 3.50GHz, 64GB RAM, and Quadro RTX 6000 GPU.

Note that only considering 24 hours will not guarantee the same set of UAVs working for consecutive days, but only ensure that the involved UAVs have sufficient battery residue to go charging in the next day. Once it is verified that a given set of UAVs can be sustainable for a whole day, two sets of UAVs can serve in alternative days to achieve full sustainability.

B. Simulation Results

The major dynamics of the environment are presented in Fig. 2 first. The solar radiation is concentrated between 7am-5pm in a day. The UAV charging rate above clouds is positively related to the solar radiation. In Subfig. 2(b), there are more users requesting services in the late morning and afternoon, which is consistent with daily working hours. To satisfy a minimum 85% of user service rate per hour, the red bars show the minimum number of UAVs needed for serving.

Parameters	Values
UAV levels (H_g, H_s, H_c)	(0,300,1400)m
Max. solar radiation intensity above clouds I_{max}	2000W/m ²
Solar radiation intensity threshold K_c	150W/m ²
UAV charging efficiency η_c	0.25
UAV maximum speeds (v_{lv}, v_{up}, v_{dn})	(6m/s, 4m/s, 4m/s)
UAV rotor disk radius r_d	0.3m
UAV charging panel area A_{Chg}	1m ²
UAV weight and air density (W, ρ)	(5x9.8N, 1.225kg/m ³)
UAV Battery Capacity E_{cap}	600Wh [15]
UAV static operational power P_{static}	5W
UAV drag profile (C_{D0}, σ, v_T)	(5e-4, 0.056, 150m/s)
Minimum user service rate p_{min}	0.85

TABLE I: Summary of Main Environment Parameters

Parameters	Values
Actor and critic networks	2 hidden layers, each with 400 hidden nodes
Neural network learning rates	0.0001 for critic and actor
Activation function	$\tanh$ (actor output), Relu (all remaining)
Regularization	L2 with $\lambda = 0.0001$
Gradient threshold	1
Smooth factor for target networks	0.001
Update frequency for target networks	1
Mini-Batch size	512
Action noise ($\sigma_{max}^2, \text{decay}, \sigma_{min}^2$)	(1.5, 0.0001, 0.2)
Experience buffer capacity	10 ⁶
Discount factor γ	0.99
Max. steps per episode	24
Max. episodes simulated	10 ⁵

TABLE II: Reinforcement Learning Parameters

Given the above dynamics and the constraints of sustainability and user demand, the convergence of episodic rewards under our designed DRL algorithms are provided in Fig. 3. It can be seen that for the same number of UAVs, convergence is almost the same for different reward coefficients, yet a larger UAV set will lead to longer convergence time due to increased dimensions of the state-action space.

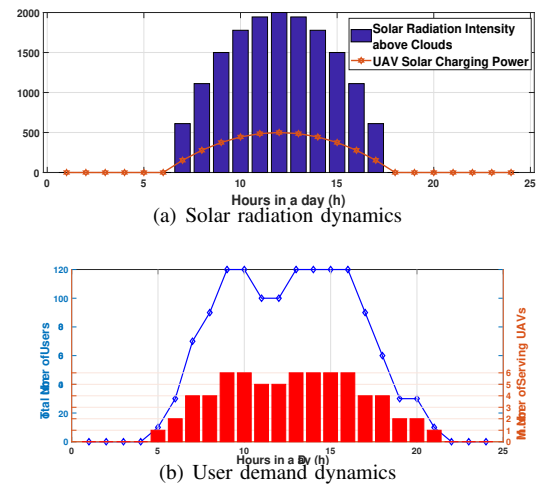


Fig. 2: Dynamics of Solar Radiation and User Demand in a Day.

Fig. 4 reveals details of the achieved optimal charging

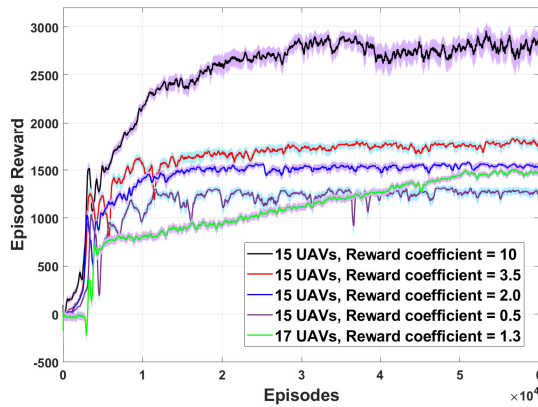


Fig. 3: Convergence of episodic reward for different numbers of UAVs and reward coefficients (c_1, c_2). The episode rewards are averaged over a window size of 300 with 95% credit interval.

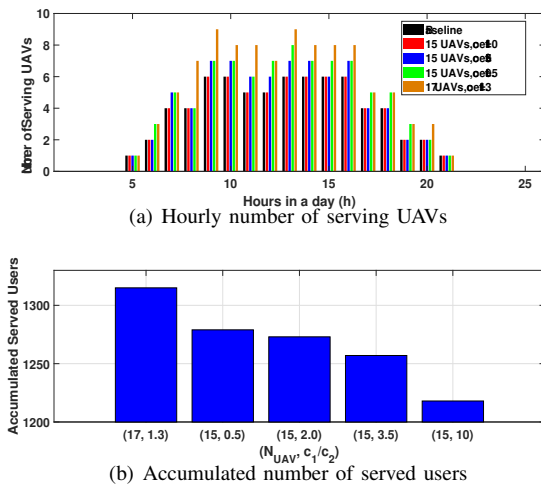


Fig. 4: Performance measures of the proposed algorithm under different parameters.

profiles in terms of hourly number of serving UAVs and the accumulated number of served users in one day. The baseline in Subfig. 4(a) gives the minimum required number of serving UAVs in each hour to satisfy the 85% user service rate. It can be observed that with smaller reward coefficients (c_1, c_2), the number of serving UAVs in each hour tend to increase. The reason is that smaller reward coefficients will result in a smaller weight C in optimization P_1 . Thus the RL agent tends to dispatch more UAVs to serve more users to get more rewards rather than make UAVs go charging or idle. When there are more UAVs available (e.g., 17 UAVs), more UAVs can serve in each hour when c_1 and c_2 are relatively low. More serving UAVs in each hour will consequently bring more served users, which is confirmed by Subfig. 4(b).

VI. CONCLUSIONS

In this paper, optimal solar charging problem has been studied in a sustainable UAV communication network, considering

the dynamic solar radiation and user service demand. The problem has been formulated into a time-coupled optimization problem and further decoupled into two sub-problems. DRL algorithms have been designed to make the sub-problems tractable. Simulation results have demonstrated the efficacy of the designed algorithms in optimally trading off the communication performance against the net energy loss.

VII. ACKNOWLEDGEMENT

This work was supported by the National Key Research and Development Program of China (2020YFB1807700), and the National Natural Science Foundation of China (NSFC) under Grant No. 62071356.

REFERENCES

- [1] Y. Zhou, N. Cheng, N. Lu, and X. S. Shen, "Multi-uav-aided networks: Aerial-ground cooperative vehicular networking architecture," *IEEE Vehicular Technology Magazine*, vol. 10, no. 4, pp. 36–44, 2015.
- [2] M. Qin, W. Wu, Q. Yang, R. Zhang, N. Cheng, H. Zhou, R. R. Rao, and X. Shen, "Green-oriented dynamic resource-on-demand strategy for multi-rat wireless networks powered by heterogeneous energy sources," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5547–5560, 2020.
- [3] R. Zhang, M. Wang, and L. X. Cai, "SREC: Proactive self-remedy of energy-constrained UAV-based networks via deep reinforcement learning," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.
- [4] R. Zhang, M. Wang, L. X. Cai, and X. Shen, "Learning to be proactive: Self-regulation of uav based networks with uav and user dynamics," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4406–4419, 2021.
- [5] Y. Sun, D. Xu, D. W. K. Ng, L. Dai, and R. Schober, "Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4281–4298, 2019.
- [6] S. Sekander, H. Tabassum, and E. Hossain, "Statistical performance modeling of solar and wind-powered UAV communications," *IEEE Transactions on Mobile Computing*, vol. 20, no. 8, pp. 2686–2700, 2020.
- [7] J. Zhang, M. Lou, L. Xiang, and L. Hu, "Power cognition: Enabling intelligent energy harvesting and resource allocation for solar-powered UAVs," *Future Generation Computer Systems*, vol. 110, pp. 658–664, 2020.
- [8] S. Khairy, P. Balaprakash, L. X. Cai, and Y. Cheng, "Constrained deep reinforcement learning for energy sustainable multi-UAV based random access IoT networks with NOMA," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1101–1115, 2020.
- [9] E. Turgut, M. C. Gursoy, and I. Guvenc, "Energy harvesting in unmanned aerial vehicle networks with 3D antenna radiation patterns," *IEEE Transactions on Green Communications and Networking*, vol. 4, no. 4, pp. 1149–1164, 2020.
- [10] A. Kokhanovsky, "Optical properties of terrestrial clouds," *Earth-Science Reviews*, vol. 64, no. 3–4, pp. 189–241, 2004.
- [11] R. Zhang, M. Wang, X. Shen, and L.-I. Xie, "Probabilistic analysis on QoS provisioning for Internet of Things in LTE-A heterogeneous networks with partial spectrum usage," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 354–365, 2015.
- [12] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *2014 IEEE global communications conference*. IEEE, 2014, pp. 2898–2904.
- [13] "Small cells and health," Available at https://www.gsma.com/publicpolicy/wp-content/uploads/2015/03/SmallCellForum_2015_small-cells_and_health_brochure.pdf, 2015.
- [14] G. Dulac-Arnold, R. Evans, H. van Hasselt, P. Sunehag, T. Lillicrap, J. Hunt, T. Mann, T. Weber, T. Degris, and B. Coppin, "Deep reinforcement learning in large discrete action spaces," *arXiv preprint arXiv:1512.07679*, 2015.
- [15] "High power density light weight drone solid state lithium battery," Available at <https://unmannedrc.com/products/high-power-density-light-weight-drone-solid-state-lithium-battery>.