Intelligent Surface-Assisted UAV Networks: A DRL Approach to Energy Efficiency

Kimchheang Chhea, Sothearath Meng
Department of Intelligent Energy and Industry
Chung-Ang University
Seoul, South Korea
chheangkim@cau.ac.kr, sothearath@cau.ac.kr

Jung-Ryun Lee *(Senior Member IEEE)*
School of Electrical and Electronics Engineering
Department of Intelligent Energy and Industry
Chung-Ang University
Seoul, South Korea
jrlee@cau.ac.kr

***Abstract*—Lower production costs have inspired studies on unmanned aerial vehicles (UAV) for wireless communication. However, limited transmission power and size of the UAV make it challenging to use advanced communication models while meeting the growing need for high data rates and energy efficiency (EE). In this paper, we study an energy-efficient UAV network enhanced by an intelligent reflecting surface (IRS) with simultaneous wireless information and power transfer (SWIPT), where the IRS is employed to improve the EE of ground user equipment (GUE). The goal is to maximize the average EE by jointly controlling the UAV's flying route, IRS phase steer, UAV transmission power, and power splitting (PS) ratio of the energy transfer technology. The formulated problem of maximizing the average EE is non-convex and thus challenging to be solved. To address this problem, we propose a deep reinforcement learning (DRL) approach. The modified reward function is implemented to enhance the efficiency of the DRL agent, which is formulated based on the expected signal-to-interference-plus-noise ratio (SINR) map. Simulation results demonstrate that the proposed DRL algorithm achieves lower energy consumption, higher data rate, and improved EE compared to the comparison algorithm.**

***Index Terms*—Intelligent reflecting surface, unmanned aerial vehicle, deep reinforcement learning, energy transfer.**

I. Introduction

UAV-aided data delivery is an important communication technology in Internet of Things (IoT) environments. This is because it can reduce the energy consumption of IoT devices by positioning the UAV near low-battery devices, thereby prolonging the network lifetime. As opposed to the ground base station (BS) that is always supplied with reliable energy source, UAVs cannot obtain energy while flying. Additionally, the location of the UAV can impact the energy consumption of both the ground device and the UAV itself. Therefore, careful trajectory planning and transmission power control are required to satisfy the increasing demands for high data rate and energy efficiency (EE) [2].

Recently, intelligent reflecting surface (IRS) has emerged as a highly promising and innovative communication paradigm within the area of wireless networks. Using IRS in UAV networks can extensively improve the communication range of the UAV without using a significant amount of energy. In addition, IRS enables passive beamforming, which mitigates high RF signal attenuation and establishes an effective transmission beam to ground devices. These capabilities to manage the wireless environment offer distinct advantages in addressing various challenges in wireless communications, such as improving spectrum efficiency and EE [4].

On the other hand, IoT networks primarily consist of wireless nodes that are geographically dispersed or spatially spread out, such as sensor nodes and device-to-device communications. One challenging issue for these networks is lowering energy usage and prolonging the lifespan of the network. Because battery replacement or regular recharging can be costly and inconvenient, *harvesting* energy from the surrounding environment is regarded as a sustainable way to offset the energy consumption of the devices [5]. Specifically, UAV-enabled wireless energy transfer holds great potential as it offers the flexibility to efficiently cover a specific area by dynamically adjusting source-to-destination distance. This adaptability allows for meeting the energy requirements of diverse nodes and enhancing energy harvesting efficiency. Furthermore, it is noted that the EE of an IoT network can be further enhanced by integration of wireless power transfer (WPT) technology into the IRS-aided UAV network platform.

Our optimization approach introduces the SINR map, a key measure of communication quality and energy efficiency, specifically defined for the DRL reward function. Integrating the SINR map with learning-based algorithm such as DRL allows for smoother training, increased stability and better adaptation to complex environment, which we will describe in more detail in Section III.

In this work, we focus on enhancing the average EE of IRS-assisted UAV networks within a simultaneous wireless information and power transfer (SWIPT) framework. Our work aims to maximize the average EE of the GUEs by simultaneously optimizing the UAV flying route, transmission power, IRS phase steer, and power splitting (PS) ratio in a IRS-aided UAV WPT network. The key contributions of this paper are outlined as follows:

- We construct a system model of the IRS-aided UAV communication system equipped with SWIPT functionality

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00251105), and by the Human Resources Development (No.20214000000280) of the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government Ministry of Trade, Industry and Energy

for IoT network, where a UAV can transfer energy to IoT devices and offer services at the same time. From the channel model for the UAV with an IRS and the energy model of the GUE, we develop an optimization problem of maximizing the average EE of the IRS-assisted UAV WPT network with decision variables of the UAV flying route, transmission power, IRS phase steer, and energy harvesting ratio of SWIPT functionality.

- To solve this optimization problem with very high computational complexity, we propose a deep reinforcement learning (DRL) algorithm. The proposed DRL algorithm introduces the concept of the SINR map (the average SINR of the UAV over the GUEs in the given network). From the SINR map, we construct the reward function using bivariate normal distribution in the proposed DRL algorithm.
- It is verified that the proposed DRL algorithm enhances the performance of the UAV in that it consumes less energy on average and maintains high data rate compared to the comparison schemes. Also, the results verify that using a UAV equipped with IRS functionality can significantly reduces the energy consumption of nodes in a network.

It is noted that our study is distinguished from previous studies on the UAV-IRS communication system [6]–[8] in that it combines energy harvesting functionality of SWIPT with IRS-aided UAV communications to increase the EE of the network system, and apply the DRL algorithm to address the non-convexity in the proposed optimization problem with low computational complexity.

II. System Architecture and Problem Formulation

A. System Architecture

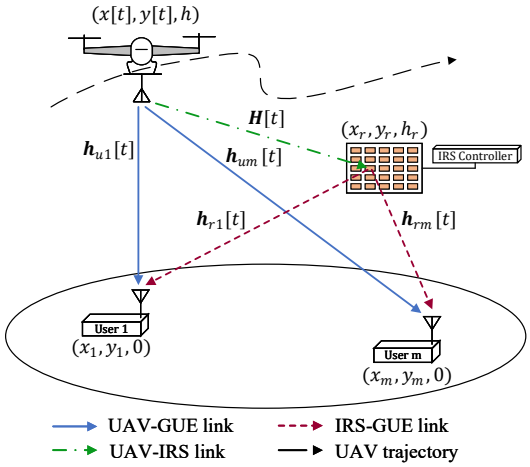

Figure 1. System model of the IRS-aided UAV communication network with energy transfer

We consider a single-antenna aerial-BS UAV serving a set of $m = \{1, 2, ..., M\}$ users, providing downlink communications within a considered area, as shown in Fig. 1. A GUE receives information and energy at the same instant due to the embedded SWIPT technology. In our assumption, we utilize a 3D Cartesian coordinate such that the location of the GUE m is fixed at $q_m = [x_m, y_m, 0]^T$. The initial flying location of the UAV is $q_{ui} = [x_0, y_0, h_0]^T$. We deploy one single IRS with multiple steerable elements which it is mounted fixed at $q_r = [x_r, y_r, h_r]^T$. The IRS consists of $N_r \times N_c$ passive reflecting element (PRE), which are consistently arranged as a uniform planar array (UPA), with N_r and N_c be the number of IRS unit in row and column, respectively. The UPA is structured such that each column contains PREs that are equidistant from each other, with a separation of s_c meters. Similarly, the UPA is composed of PREs arranged in rows that are equidistant, with a spacing of s_r meters. The UPA allows each PRE to independently re-scatter the incoming signal, and this process is characterized by a reflection coefficient comprising an amplitude a ranging from 0 to 1 and a phase steer $\theta_{n_r,n_c} \in [0, 2\pi]$, i.e. $r_{n_r,n_c} = ae^{j(\theta_{n_r,n_c})}$, $\forall n_r \in \{1, 2, \ldots, N_r\}$, and $\forall n_c \in \{1, 2, \ldots, N_c\}$. In this paper we use fixed $a = 1$, and phase steer θ_{n_r,n_c} can be modified by the IRS decision maker.

B. Channel Model for UAV data delivery with IRS

The UAV is dispatched to provide services to all GUEs. Unlike traditional ground-based networks, where Rayleigh fading is commonly employed for small-scale fading, Rician fading is deemed more suitable for UAV-ground communications. This choice is justified by the typically prevalent Line-of-Sight (LoS) channel component and the occurrence of local scattering in UAV-ground communication scenarios. Thus, we utilize the channel model of Rician fading for both the UAV-GUE link and the IRS-GUE link. By considering the substantial signal attenuation and loss in reflection, we assume that signals reflected by the IRS two or more times have minimal power and are consequently disregarded. The channel between the UAV and the GUE m can be described as

$$\mathbf{h}_{um}[t] = \sqrt{\frac{\beta_0}{d_{um}^{\alpha_{um}}[t]}} \left(\sqrt{\frac{\kappa}{\kappa+1}} + \sqrt{\frac{1}{1+\kappa}} \tilde{\mathbf{h}}_{um}[t] \right), \quad (1)$$

here, α_{um} is the path loss factor specifically associated with the link of the UAV to the GUE m. The Rician factor is denoted as κ, and $\tilde{\mathbf{h}}_{um}[t] \sim \mathcal{CN}(0, 1)$ represents the scattering element of GUE m during time slot t, following a complex circularly symmetric Gaussian distribution with zero mean and unit variance. At time t, the channel representing the communication of the UAV-IRS link is expressed as

$$\mathbf{H}[t] = \sqrt{\frac{\beta_0}{d_{ur}^2[t]}} \tilde{\mathbf{H}}[t], \quad (2)$$

where β_0 is gain at $d_0 = 1$m reference distance, and $\tilde{\mathbf{H}}[t]$ is the LoS channel of the UAV-GUE link, given in (7). We

denote $\theta_{ur}[t]$, $\zeta_{ur}[t]$ and z as the angle-of-arrivals at the IRS, and the height of the UAV, respectively, where

$$\sin\theta_{ur}[t] = \frac{z - h_r}{d_{ur}[n]}, \tag{3}$$

$$\sin\zeta_{ur}[t] = \frac{x_r - x[t]}{\sqrt{(x_r - x[t])^2 + (y_r - y[t])^2}}, \tag{4}$$

$$\cos\zeta_{ur}[t] = \frac{y[n] - y_r}{\sqrt{(x_r - x[t])^2 + (y_r - y[t])^2}}. \tag{5}$$

At time slot t, the channel model from the IRS to the GUE m is described as

$$\mathbf{h}_{rm}[t] = \sqrt{\frac{\beta_0}{d_{rm}^{\alpha_{rm}}[t]}}\left(\sqrt{\frac{\kappa}{\kappa+1}}\mathbf{h}_{rm}^{LoS}[t] + \sqrt{\frac{1}{1+\kappa}}\tilde{\mathbf{h}}_{rm}[t]\right), \tag{6}$$

where $\tilde{\mathbf{h}}_{rm}[t] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_rN_c})$, $\mathbf{I}_{N_rN_c}$ is the covariance matrix, and $\mathbf{h}_{rm}^{LoS}[t]$ is given by (8). The reflector coefficient of the IRS at time slot t is described by

$$\mathbf{\Theta}[t] = \text{diag}(\theta[t]) \in \mathbb{C}^{N_rN_c \times N_rN_c}, \tag{9}$$

where $\theta[t] = [e^{j\theta_{1,1}}[t], ..., e^{j\theta_{n_r,n_c}[t]}, ..., e^{j\theta_{N_r,N_c}[t]}]^T \in \mathbb{C}^{N_rN_c \times 1}$. The composite of the UAV-GUE channel can be formulated as

$$\mathbf{h}_m^H[t] \triangleq \mathbf{h}_{rm}^H[t]\mathbf{\Theta}[t]\mathbf{H}[t] + \mathbf{h}_{um}^H[t]. \tag{10}$$

Because of the scattering elements present in the Rician fading channels described in equations (1) and (6), the overall channel becomes probabilistic in nature. To ensure effective control over the IRS phase for coherent signal composition at the GUE and to facilitate UAV route planning, it is necessary to approximate and provide real-time channel information between the PRE, UAV, and the individual GUE. This information is essential to be shared with the central decision maker of both the IRS and the UAV during the entire flight operation.

C. Energy Model for IRS-aided UAV data delivery

In our work, we consider the SWIPT-equipped GUE using PS technique. With the PS ratio ρ_m, the GUE can exploit energy and information signal at the same instant with the received SINR at the GUE m, which can be written as

$$SINR_m[t] = \frac{\rho_m p_m[t]|\mathbf{h}_m^H[t]|^2}{\rho_m \sum_{m' \neq m} p_{m'}[t]|\mathbf{h}_{m'}^H[t]|^2 + \sigma_m^2}, \tag{11}$$

where p_m and σ_m are the received power and noise, respectively, at the GUE m. The energy dissipation of the GUE m can be described as

$$\begin{aligned} ED_m[t] =& P_c + p_m - (1-\rho_m)\big(p_m[t]|\mathbf{h}_m^H[t]|^2 \\ &+ \sum_{m' \neq m} p_{m'}[t]|\mathbf{h}_{m'}^H[t]|^2\big), \end{aligned} \tag{12}$$

where P_c is the circuit power consumption and $(1-\rho_m)p_m[t]|\mathbf{h}_m^H[t]|^2$ is the energy signal received at the GUE m. Therefore, we can express the data rate received at the GUE m as

$$R_m[t] = \log(1 + SINR_m[t]). \tag{13}$$

D. Problem Formulation

From (11) and (12), we define the EE at the GUE m as

$$EE_m(q_u, \rho_m, p_m, \mathbf{\Theta})[t] = \frac{R_m(q_u, \rho_m, p_m, \mathbf{\Theta})[t]}{ED_m(q_u, \rho_m, p_m, \mathbf{\Theta})[t]}. \tag{14}$$

Accordingly, we formulate a problem that simultaneously optimizes the route planning of the UAV, PS ratio, transmission power, and IRS phase steer to maximize the average EE which can be described as

$$\max_{q_u, \rho_m, p_m, \mathbf{\Theta}} \frac{1}{M}\sum_{m=1}^{M} EE_m(q_u, \rho_m, p_m, \mathbf{\Theta})[t] \tag{15}$$

$$\text{s.t.} \quad C1: p_m[t] \leq p_{max}, \tag{16}$$

$$C2: 0 \leq \Theta[t] \leq 2\pi, \tag{17}$$

$$C3: 0 < \rho_m[t] \leq 1, \tag{18}$$

$$C4: \|q_u[t+1] - q_u[t]\| \leq D_{max}, \ \forall t, \tag{19}$$

where constraints $C1, C2, C3$ indicate that the power cannot exceed the maximum power, phase steer is within the range of $[0, 2\pi]$, and the PS ratio is in range of $[0, 1]$, respectively. In addition, constraint $C4$ defines the displacement of the UAV from one location at time slot t to the next location at time slot $t+1$. The optimization problem becomes non-convex since all constraints exhibit non-linear relationships with the control variables. In our work, we handle high-dimensional control parameters such as the route planning of the UAV, PS ratio, transmission power allocation, and IRS phase steer in dynamic environment. This justifies the use of DRL with reduced computational complexity.

$$\begin{aligned} \tilde{\mathbf{H}}[t] =& \Big[1, e^{-j2\pi d\frac{\sin\theta_{ur}[t]\cos\zeta^{ur}[t]}{\lambda}}, ..., e^{-j2\pi d(N_r-1)\frac{\sin\theta_{ur}[t]\cos\zeta^{ur}[t]}{\lambda}}\Big]^T \\ &\times \Big[1, e^{-j2\pi d\frac{\sin\theta_{ur}[t]\sin\zeta^{ur}[t]}{\lambda}}, ..., e^{-j2\pi d(N_c-1)\frac{\sin\theta_{ur}[t]\sin\zeta^{ur}[t]}{\lambda}}\Big]^T \end{aligned} \tag{7}$$

$$\begin{aligned} \mathbf{h}_{rm}^{LoS} =& \Big[1, e^{-j2\pi d\frac{\sin\theta_{rm}[t]\cos\zeta^{rm}[t]}{\lambda}}, ..., e^{-j2\pi d(N_r-1)\frac{\sin\theta_{rm}[t]\cos\zeta^{rm}[t]}{\lambda}}\Big]^T \\ &\times \Big[1, e^{-j2\pi d\frac{\sin\theta_{rm}[t]\sin\zeta^{rm}[t]}{\lambda}}, ..., e^{-j2\pi d(N_c-1)\frac{\sin\theta_{rm}[t]\sin\zeta^{rm}[t]}{\lambda}}\Big]^T \end{aligned} \tag{8}$$

III. Proposed Algorithm

Reinforcement learning agents utilize Q-learning to discover effective strategies through environmental exploration. Each decision yields rewards, and the agent strives to develop a strategy that yields the highest long-term value when considering future discounted rewards. We represent the expected cumulative reward as $Q^{\pi}(s,a)$, which evaluates how beneficial it is to take action a at state s_t while following strategy π. The core challenge is identifying the supreme action values, denoted as $Q^{*}(s,a) = \max_{\pi} Q^{\pi}(s,a)$, which lead to optimal decision-making. To track these learned values, the algorithm maintains a lookup matrix called the Q-table that maps each state-action combination to its estimated worth.

At each time slot t, The Q-value is computed based on the present state and the action chosen in the previous step. The recorded value is kept in a Q-function, which plays a crucial role in determining the policy Π. Through DRL, the agent learns to make decisions that optimize long-term aggregate benefits instead of just pursuing short-term gains.

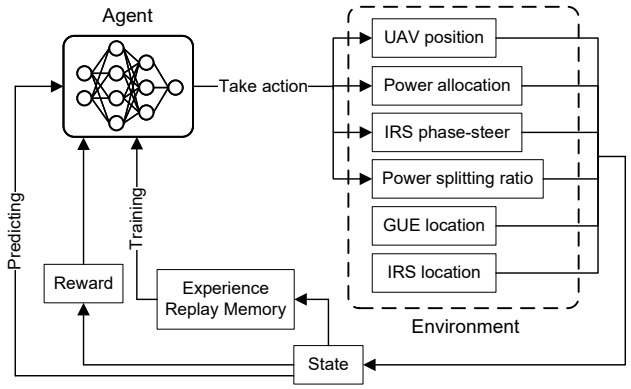

Figure 2. Proposed DRL model

During each time slot t, the Q-value and Q-function are continuously updated based on the current state, previous actions, and received reward to enhance the agent's decision-making. This update process is performed using the equation as follows

$$Q_{t+1}(s_t,a_t) \leftarrow Q_t(s_t,a_t) + \eta[r_t + \gamma \max_{a} Q_t(s_{t+1},a)] \quad (20)$$

where η and γ are the step size and discount rate, respectively. In equation (20), the reward r_t is acquired from $r := S \times A \rightarrow R$, where $\mathbb{E}\{r_t|(s,a,s^{'}) = (s_t,a_t,s_{t+1})\} = R^{s^{'}}_{s,a}$. From iterative updating of the equation, the optimal value function can be acquired as follows

$$Q^{*}(s,a) = \mathbb{E}_{s'}[R + \gamma \max_{a^{'}} Q^{*}(s^{'}a^{'})|s,a]. \quad (21)$$

In the DRL based model, we make the assumption that a single decision maker, acting as an agent, governs both the IRS and the UAV. The agent receives information on state

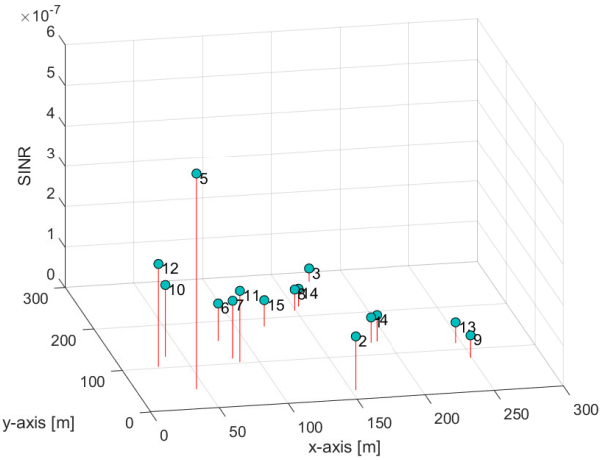

Figure 3. SINR levels of the UAV (located at $[0,0,z]^T$) to given 15 GUEs

s_t from the state space $\mathbf{S}$ during time slot t, which includes the positions of the UAV and all GUEs, and the IRS phase steer. The agent acquires the current state and make choice according to the decision policy Π, then selects one action a_t from a set of possible action $\mathbf{A}$ that includes the direction of movement for the UAV, PS ratio, transmission power, and IRS phase steer. Following the agent's action, it receives a reward or penalty r_t determined by the average EE of the UAV. We explain the detailed definition of the states, actions, and reward function in our work, as follows.

1) States: The state space of the proposed DRL model is defined by

$$s^t = \{q_u[t], \rho[t], p[t], \mathbf{\Theta}[t], \mathbf{h}^H_m[t], t\} \quad (22)$$

where $q_u[t] = [x[t], y[t], h]^T$ is the UAV two-dimensional (2D) coordinate, $\rho[t] \in [0,1]$ is the PS ratio, $p[t]$ is the transmission power, $\mathbf{\Theta}[t] \in \mathbb{C}^{N_r N_c \times N_r N_c}$ is the IRS phase steer at time slot t.

2) Actions: The action space of the proposed DRL model is designed as

$$a^t = \{\Delta q_u[t], \Delta\rho[t], \Delta p[t], \Delta\mathbf{\Theta}[t]\}, \quad (23)$$

where $\Delta q_u[t] \in \{(-\delta x,0),(\delta x,0),(0,-\delta y),(0,\delta y),(-\delta xy,0),(\delta xy,0),(-\delta yx,0),(\delta yx,0),(0,0)\}$ means the moving directions of the UAV, $\Delta\rho[t] \in \{0,...,1\}$, $\Delta p[t] \in \{0,...,p_{max}\}$, $\Delta\mathbf{\Theta}[t] \in \{0,...,2\pi\}$.

3) Reward: Using the concept of the SINR map, we build a reward function, which indicates the average EE of the UAV under given network topology. The reward function is designed as

$$r(\tilde{q}_u, q_u, \rho_m, p_m, \mathbf{\Theta})[t] = \frac{1}{M}\sum_{m=1}^{M} A_m f_{\tilde{q}_m}(\tilde{q}_u) EE_m(q_u, \rho_m, p_m, \mathbf{\Theta})[t], \quad (24)$$

where f is the multivariate (bivariate in our work) normal distribution, $\tilde{q}_m$ and $\tilde{q}_u$ denote the projections of q_m and q_u

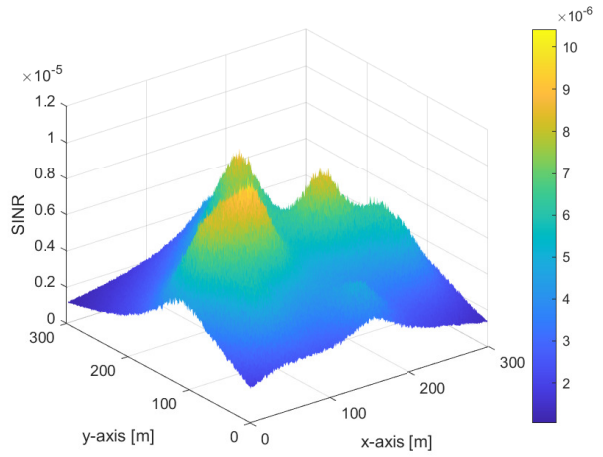

Figure 4. Average SINR values of the UAV over given 15 GUEs

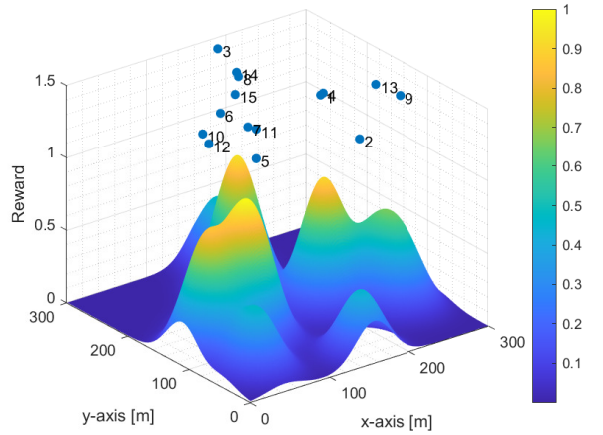

Figure 5. Reward values of the UAV in 3D

over x-y plane, respectively, and A_m is the normalization weight. It is noted that the function f is multiplied by EE_m to obtain a continuous and differentiable reward function while preserving the characteristics of the SINR distribution of multi-modal Gaussian distribution. f is given by

$$f_{\vec{\psi}}(\vec{a}) = \frac{\exp\left(-\frac{1}{2}(\vec{a}-\vec{\psi})^{\mathrm{T}}\mathbf{\Delta}^{-1}(\vec{a}-\vec{\psi})\right)}{\sqrt{(2\pi)^k|\mathbf{\Delta}|}}, \tag{25}$$

where $\vec{a}$ is a column vector, $\vec{\mu}$ is the mean vector, k is the dimension of the function, $\mathbf{\Delta}$ is the covariance matrix and $|\mathbf{\Delta}| \equiv \det \mathbf{\Delta}$ is the determinant of the $\mathbf{\Delta}$.

Figs. 5 shows the average EE of the UAV in 3D, under the same environment in Fig. 4. It is noticeable that the EE in Fig. 5 is a surface, and the product of EE_m and the bivariate function gives the EE a smoother contour shape as opposed to the average SINR value in Fig. 4. In addition, the EE of the UAV is designed in such a way that the highest EE region provides the highest reward, which induces the UAV to fly over or hover around this region with high probability.

Consequently, the penalty of the proposed DRL algorithm is defined as

$$a_t^{'} = \begin{cases} r_t & \text{satisfies} \quad C1, C2, C3, C4, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

The objective of the deep neural network (DNN) is to reduce the value of the loss function, which can be expressed as follows

$$L(\phi) = \mathbb{E}[(y - Q(s_t, a_t, \phi))^2], \tag{27}$$

where $y = r_t + \gamma \max_{a \in A} Q_{prev}(s_t, a_t, \phi)$. During the training phase of the DNN, the parameter ϕ is updated using a technique called experience replay. This involves randomly selecting a minibatch, denoted as $\hat{D}$, from the experience replay memory D. The selected minibatch is then utilized as the input data for updating the parameter ϕ of the neural network. This approach allows for efficient and effective utilization of past experiences to improve the training process of the DNN.

Error gradient is obtained by chain rule, which is given by

$$\nabla_\phi L \approx \frac{1}{|\hat{D}|} \sum 2(y - Q(s_t, a_t, \phi)) \nabla_\phi Q(s_t, a_t, \phi). \tag{28}$$

At each iteration, the agent modifies its decision-making strategy based on the current estimate of the Q-value. The agent employs an ϵ-greedy policy to choose an action from the action space. This policy is defined as follows

$$a_t^{'} = \begin{cases} \text{argmax}_{a \in A} Q(s_t, a_t, \phi) & \text{with probability } 1 - \epsilon \\ \text{random action} & \text{with probability } \epsilon. \end{cases} \tag{29}$$

IV. Results and Discussion

Table I
Simulation Parameters.

Parameter	Value
Coverage area	300m×300m
Number of the GUEs	15
Number of reflecting units	{10,20,...,80}
Velocity of the UAV	5m/s
UAV flying height	100m
IRS's height	30m
transmission power	{10,12,14,...,28} [dBm]
Energy transfer efficiency	50%
Path loss exponent (NLoS)	3.6
Path loss exponent (LoS)	2.2
Rician factor	2
Discount factor	0.8

In this section, we evaluate the effectiveness of the proposed DRL with SINR map-based reward through comprehensive evaluations of various simulations. For comparison, we select the Successive Fly-and-Hover (SHF) scheme with random IRS phase steer, DRL without IRS and REINFORCE. REINFORCE is derived as a Monte-Carlo policy gradient learning algorithm, which trains the agent to generate a stochastic policy. Due to the challenge of balancing exploration and exploitation during training, REINFORCE often converges to

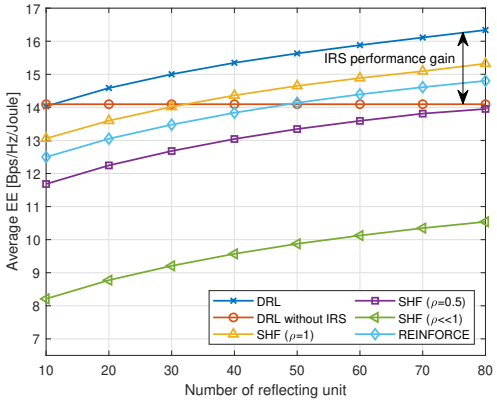

Figure 6. Average EE vs. the number of reflecting unit

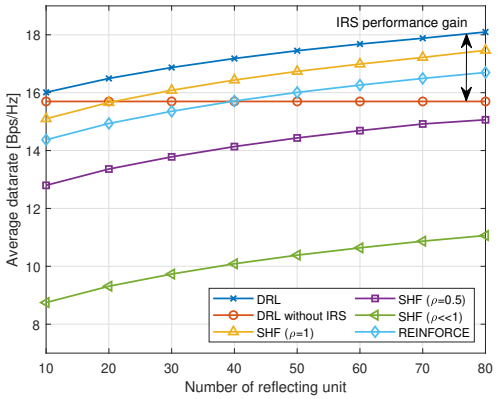

Figure 7. Average data rate vs. the number of reflecting unit

suboptimal solution. The Table I provides a summary of the parameters employed in the simulation.

In the following figures, we analyze the performances with various number of reflecting units. In Fig. 6 the average EE increases as the number of reflecting unit goes up. We can clearly see the IRS performance gain of the proposed algorithm compared to the DRL without IRS method.

Fig. 7 shows the data rate achieved as we increase the number of reflecting units. It is verified that the proposed method achieves a higher data rate compared to that of the SHF with $\rho = 1$. Although REINFORCE achieves lower energy consumption than SHF with $\rho = 1$, its datarate is also lower than the SHF, which results in lower EE. Fig. 8 depicts the performance considering the average energy consumption versus the number of reflecting units. From the result in Fig. 8, we can verify that the proposed algorithm takes advantage of the reflecting units of the IRS to reduce the energy consumption. Furthermore, as the number of reflecting units increases to a certain amount, the performance gradually goes to convergent state, highlighting the limitation of IRS. Additionally, the results show that implementing the proposed DRL with 10 IRS elements performs similarly to the baseline version without any IRS components. Therefore, it is important

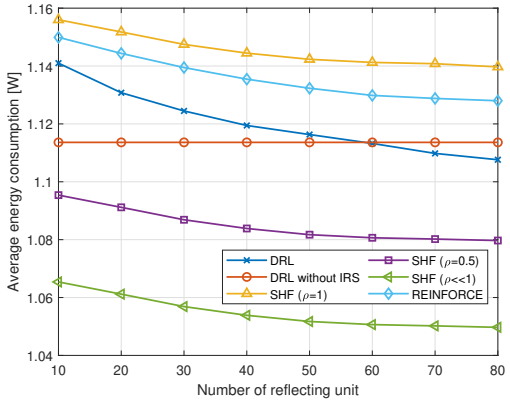

Figure 8. Average energy consumption of the GUE vs. the number of reflecting unit

to employ more than 10 IRS units to reap the advantage of the IRS.

V. Conclusion

In this paper, we analyze an IRS-aided UAV WPT network with SWIPT, formulating an average energy efficiency maximization problem. To tackle this, we use a DRL approach and propose a reward function based on the EE of GUEs to jointly optimize the UAV's flight path, IRS phase steering, transmission power, and power splitting ratio. Simulation results highlight the benefits of our algorithm, both with and without IRS. Future work will explore EE for 3D UAV route planning in a multi-UAV environment using multi-agent DRL.

References

[1] A. Fotouhi et al., "Survey on UAV Cellular Communications: Practical Aspects, Standardization Advancements, Regulation, and Security Challenges," in IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3417-3442, Fourthquarter 2019, doi: 10.1109/COMST.2019.2906228.

[2] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," Proc. IEEE, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.

[3] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," IEEE Commun. Mag., vol. 58, no. 1, pp. 106–112, Jan. 2020.

[4] Y. Liu et al., "Reconfigurable Intelligent Surfaces: Principles and Opportunities," in IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1546-1577, thirdquarter 2021, doi: 10.1109/COMST.2021.3077737.

[5] N. Zhao, S. Zhang, F. R. Yu, Y. Chen, A. Nallanathan and V. C. M. Leung, "Exploiting Interference for Energy Harvesting: A Survey, Research Issues, and Challenges," in IEEE Access, vol. 5, pp. 10403-10421, 2017, doi: 10.1109/ACCESS.2017.2705638.

[6] Z. Wei et al., "Sum-Rate Maximization for IRS-aided UAV OFDMA Communication Systems," in IEEE Transactions on Wireless Communications, vol. 20, no. 4, pp. 2530-2550, April 2021, doi: 10.1109/TWC.2020.3042977.

[7] Y. Pan, K. Wang, C. Pan, H. Zhu and J. Wang, "UAV-Assisted and Intelligent Reflecting Surfaces-Supported Terahertz Communications," in IEEE Wireless Communications Letters, vol. 10, no. 6, pp. 1256-1260, June 2021, doi: 10.1109/LWC.2021.3063365.

[8] Z. Li, W. Chen, H. Cao, H. Tang, K. Wang and J. Li, "Joint Communication and flying route Design for Intelligent Reflecting Surface Empowered UAV SWIPT Networks," in IEEE Transactions on Vehicular Technology, 2022, doi: 10.1109/TVT.2022.3196039.