Int. J. Electron. Commun. (AEÜ) 171 (2023) 154933

Contents lists available at ScienceDirect

International Journal of Electronics and Communications

journal homepage: www.elsevier.com/locate/aeue

Regular paper

Enhancing UAV communication links with Reconfigurable intelligent surfaces

Nameer Mufeed Salih [a], Mahmoud Aldababsa [b,*], Khalid Yahya [a]

[a] Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul 34[illegible] Turkey
[b] Department of Electrical and Electronics Engineering, Nisantasi University, Istanbul 34467, Turkey

ARTICLE INFO

Keywords:
Reconfigurable intelligent surface (RIS)
unmanned aerial vehicle (UAV)
simultaneous wireless information and power transfer (SWIPT)
deep Q-network (DQN)
deep deterministic policy gradient (DDPG)

ABSTRACT

Reconfigurable intelligent surfaces (RIS) have emerged as a prominent and widely debated solution to enhance the energy efficiency of wireless communications. This article explores the potential of combining RIS with unmanned aerial vehicles (UAVs)-RIS to provide on-demand deployment services in dynamic environments. However, the energy limitations of battery-powered UAVs can curtail the advantages of UAV-RIS systems. To address this challenge and enhance the durability of UAV-RIS deployments, we introduce an energy harvesting technique for simultaneous wireless information and power transfer (SWIPT) coupled with optimized resource allocation and energy harvesting from incoming radio frequency (RF) signals. In contrast to previous research, our approach involves the division of reflecting arrays across geometric space, facilitating simultaneous information transfer and energy harvesting. Additionally, we are developing deep Q-network (DQN) and deep deterministic policy gradient (DDPG) techniques to dynamically allocate UAV-RIS resources in both temporal and spatial dimensions. This allocation maximizes the overall harvested energy while upholding communication quality for every user. Simulation results conclusively demonstrate the substantial superiority of the proposed UAV-RIS SWIPT system over the benchmark.

1. Introduction

In recent years, reconfigurable intelligent surfaces (RIS) have emerged as a transformative wireless communication and networking technology. These intelligent surfaces are composed of an array of passive, programmable elements that can manipulate electromagnetic waves in wireless environments. RIS elements can adaptively alter the phase and amplitude of impinging electromagnetic signals, allowing for precise control over signal propagation. This capability opens up new horizons in wireless communications, enabling advancements in beamforming, signal enhancement, and interference mitigation [1,2].

Simultaneous wireless information and power transfer (SWIPT) has emerged as a groundbreaking paradigm in the ever-evolving landscape of wireless communications and energy harvesting. In SWIPT, the same radio-frequency signals transmit data and deliver power to receiving devices. SWIPT harnesses innovative techniques, such as energy harvesting and reconfiguration, to enable devices to concurrently acquire information and replenish their energy reserves from the ambient electromagnetic environment [3].

Unmanned aerial vehicles (UAVs) are a potential but challenging platform for investigating autonomous and cooperative control. The UAV is an electromechanical system that may operate autonomously, remotely, or both. Fixed-wing UAVs have been a prevalent type of UAV in the market, and they have been examined in numerous settings ranging from academic to commercial due to their energy-efficient performance while carrying payloads [4,5].

The use of UAVs has significantly increased in various fields, including military, transportation, and surveillance. Reliable communication links for UAVs are crucial for safe and efficient operation, but they face challenges such as interference, signal attenuation, and limited energy resources. To overcome these challenges, researchers have been exploring new technologies, including RISs, which can enhance communication links by modifying the propagation of radio waves. Additionally, optimizing the energy efficiency of UAV communication links is crucial to extend mission duration and reduce operational costs. This can be realized by utilizing the concept of SWIPT. This paper discusses using RIS to improve UAV communication links and integrating energy-efficient techniques, such as deep learning, to optimize energy consumption [6,7,8].

Deep reinforcement learning (DRL) is a powerful subset of machine

* Corresponding author.
E-mail addresses: nameermof@gmail.com (N.M. Salih), mahmoud.aldababsa@nisantasi.edu.tr (M. Aldababsa), hayahya@gelisim.edu.tr (K. Yahya).

https://doi.org/10.1016/j.aeue.2023.154933
Received 6 August 2023; Accepted 21 September 2023
Available online 25 September 2023

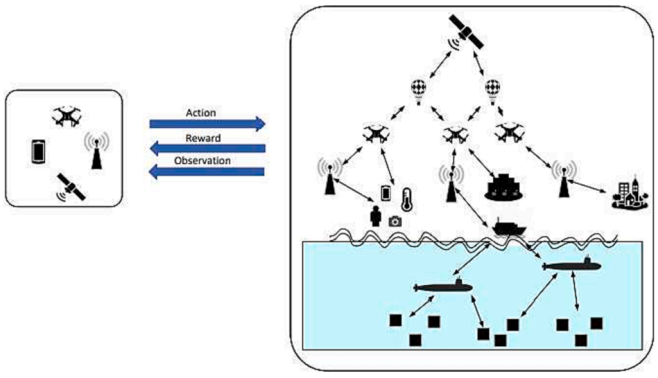

Fig. 1. RL model in wireless networks.

learning that combines deep learning and reinforcement learning principles to enable agents to make intelligent decisions in complex and dynamic environments. DRL has gained significant attention recently due to its remarkable ability to tackle various real-world problems, from autonomous robotics and game playing to optimization and recommendation systems. The DRL calculations have exhibited exceptional outcomes in handling asset-the-board difficulties in UAV-helped remote organizations [9]. To extend the association's total rate, an iterative technique with two phases of using the significant Q-learning computation and differentiation of bent estimation was made to smooth out the UAV's regions, convey beamforming, and UAV-clients connection [10]. In [11], the UAV was utilized to charge ground gear and gather information using the wireless power transfer (WPT) approach. DRL calculations have likewise been utilized in UAV-helped remote correspondence way arranging [12,13]. The authors of [14] mended a DRL calculation for recognizing the flight course, transmission power, and related cells in UAV-controlled remote organizations, g the reverberation state organization of [15]. In [16], authors propose novel approach to improve communication links by swarm UAVs with integrated RISs (SARIS). The auth suggest that th swarm of UAVs with RISs can perform dyn config of the radio wave propagation to optimize com addition, the work presented a theoret analysis simulation results to demonstrate the effectiveness he proposed framework in improving signal quality and ergy consum . The authors of [17] involved the deep -learn DQL) calculation for the UAV route in light of th ignal power luated employing a tremendous multiple-inp ultiple-output (MI method. Q-learning was utilized in [18] t ntrol th velopment of a few UAVs in static and dynamic client un an irregular walk model. The works [19,20] depicted the L tegy fo ening information bundle misfortune i helped ence ve and information-gathering framewor n [2 he auth posed a new approach to improve wireles tworks' ergy efficie y and communication performance. They s ted nd UAVs to create an optimal communication in ture for lti-user scenarios. They also presented a theoretical a is and simulation results to demonstrate the effectiveness and e efficiency of the proposed system. The authors found that using U Vs and RIS can lead to significant energy savings by reducing the energy consumption of the transmitters and amplifiers, as well as improving signal quality and coverage.

Reinforcement learning (RL) calculation is planned for specialists to settle on a grouping of choices. The specialist in Fig. 1 can be any part of a handling unit in remote organizations, like UAVs, servers, ground clients, etc. The specialists collaborate with the climate to accomplish the best prize through experimentation and learning. Subsequently, DRL approaches have given every hub independence to make the remote organization keener [22,23,24]. The remote conditions are changed into a computerized Internet of Things (IoT) area, UAV flying speed, point, position, and channel model between each organization piece. The DRL strategies are then used to prepare the specialists. We use profound DQL, DDQL, and proximal approach streamlining (PPS) procedures for the discrete issue. At the same time, for a consistent plan, we utilize profound deterministic arrangement slope (DDPG) and proximal strategy improvement (PSI). Moreover, we apply multi-specialist figuring out how to abbreviate the postponement in data transmission between the unified handling unit and clients.

This article is motivated by the rising significance of RIS in the realm of wireless communications and their potential synergy with UAVs to provide dynamic deployment services. The critical challenge addressed here is the energy limitations faced by bat d UAVs, which can restrict the efficacy of UAV-RIS syste To over this obstacle and bolster the longevity of UAV-RIS oyments, a g ndbreaking energy harvesting technique for SWIPT troduced, upled with optimized resource allocation a ergy ha ing fr coming RF signals. What sets our appro part is the im vision of passive reflected arrays acr eometri pace, en g concurrent information transfer and vesting. Furthermore, we are pioneering the use QN a PG tech s to allocate UAV-RIS resources across temporal sp dimensions dynamically. This dynamic on not only zes overall harvested energy but also en s th tenance of mmunication quality for every user. Our rigorous sim n results irrefutably affirm the remarkable super f the prop UAV-RIS SWIPT system compared to es shed benchmarks, ma ng a significant contribution to wireless c munication and UAV deployment.

2. PG and D algorithms

DDP DQN are distinct DRL algorithms that can play a vital role nabling intelligent decision-making and optimization in UAV-RIS , ultimately contributing to improved energy efficiency and communication performance. In the following, we provide their functionalities, training processes, and adaptation to the UAV-RIS.

2.1. DQN

Functionality: DQN is a foundational DRL algorithm that excels in discrete action spaces, making it suitable for tasks where actions are chosen from a finite set of possibilities. In the UAV-RIS context, DQN can be applied to discrete actions like choosing among predefined UAV maneuvers, power levels, or communication strategies.

Training Process: DQN approximates the Q-values of state-action pairs using a deep neural network. The training involves interacting with the environment, collecting experiences, and updating the Q-network through Q-learning. Experience replay breaks correlations in the data, and target networks are employed to stabilize learning.

Adaptation to UAV-RIS: To adapt DQN to the UAV-RIS context, the state representation can be customized to include variables like UAV position, RIS configurations, channel conditions, and energy levels. The action space can encompass the UAV's discrete choices, such as selecting from a predefined set of trajectories, power levels, or communication modes. DQN can be trained to optimize these actions for efficient and effective UAV-RIS interactions.

2.2. DDPG

Functionality: DDPG is a robust DRL algorithm designed for continuous action spaces, making it suitable for problems where the agent needs to select actions with a high degree of granularity, such as controlling UAVs in a UAV-RIS system. DDPG combines concepts from both policy gradients and Q-learning. It learns a deterministic policy, mapping states directly to actions, and maintains a value network to estimate the expected cumulative reward.

Training Process: DDPG employs an actor-critic architecture. The

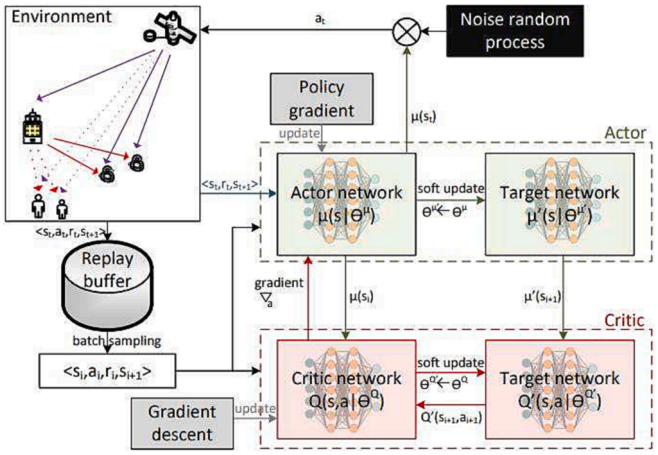

Fig. 2. Design of DQN calculation.

actor network learns the policy, while the critic network evaluates the quality of the policy. Training involves collecting experiences, computing gradients, and updating the actor and critic networks using backpropagation. Experience replay and target networks are also employed to stabilize training.

Adaptation to UAV-RIS: In the context of UAV-RIS systems, DDPG can be adapted by customizing the state representation to incorporate information relevant to the UAV's position, RIS configurations, channel conditions, and energy constraints. The action space can include parameters for UAV trajectory, power allocation, and communication strategy. DDPG can be trained to optimize these actions to achieve energy-efficient and high-performance communication wi elements.

3. Proposed system

As previously stated, enhancing decision-mak me t eet UAV assisted wireless networks' stringent standard itical. I is section, we study approaches based on the DRL algorith de trajectory and allocating resources to mize t ergy efficiency and network sum rate. For a trade-off lem, we offe rategy based on the DQL and the dueling DQL m ese approach ovide the most economical trajectory and s rate meeting severe distance, flight time, and communica s requiremen hese approaches may operate in a dynamic en ment with rando istributed and mobile users. The desig d func is adaptable to numerous rapid deployment mission sum maximization. Furthermore, we assume perfect channel s mation and static user conditions.

3.1. Dqn- a al hm for ases

In as s 2, a specialist controls the UAV's cooperation wi mate. We pect two deep neural networks (DNNs), the assessment bjective organizations. The objective organization has a similar const n as the assessment organization. However, it just updates consisten . The activity a_t is determined given the Q-values and a-covetous strategy. The award r_t is then got from the climate. It ought to be noticed that the proposed DQN-based calculation can upgrade the direction of the UAV in the limitedly discrete activity space. Thus, we characterize the activity space in every one of T_S as A, which contains the activities recorded beneath:

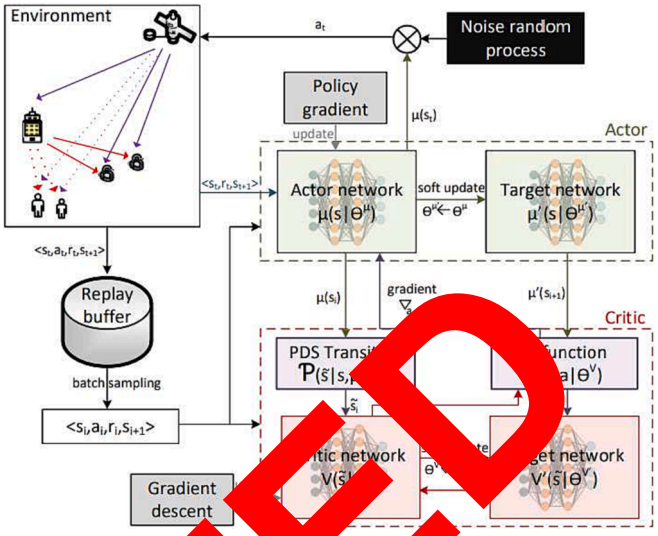

Fig. Architect the D algorithm.

$$A = \begin{cases} [x^{max},\ 0,\ 0] \\ [-x^{max},\ 0,\ 0] \\ \left[\frac{x^{max}}{2},\ 0,\ 0\right] \\ \left[\frac{-x^{max}}{2},\ 0,\ 0\right] \\ [0, y^{max},\ 0] \\ [0, -y^{max},\ 0] \\ \left[0, \frac{y^{max}}{2},\ 0\right] \\ \left[0, \frac{-y^{max}}{2},\ 0\right] \\ [0,\ 0, z^{max}] \\ [0,\ 0, -z^{max}] \\ [0,\ 0,\ 0] \end{cases} \quad (1)$$

The change comprising of s_t, a_t, r_t, s_{t+1} is then put away in an encounter replay memory. At the point when there are an adequate number of changes in the experience replay memory, the learning methodology starts. To prepare the assessment organization, a smaller than normal bunch haphazardly tests M changes, the misfortune capability for refreshing the assessment organization might be registered, which can be addressed as:

$$L_i(\delta_i) = \mathbb{E}_{s,a}[(r + \gamma.\max\{Q(s', a'|\delta_{i-1}) - Q(s, a|\delta_i)\})^2] \quad (2)$$

The algorithm is designed to optimize the communication links between a UAV and a ground station using a RIS. Here is a detailed description of the algorithm:

1. Define the environment with the state, movement, and grant. The state consists of the UAV's height, the RIS's beamforming vector, and the frequency band used for communication. The movement is defined as the velocity of the UAV and the direction of the movement. The grant is a reward function that reflects the system's energy efficiency and communication performance.

Table 1
Simulation parameters.

Parameter	Explanation	Value
K	Number of users served by an UAV	4
MUAV	Mass of the UAV	9.65 kg
I	Number of iterations	40
Tfly	Time taken for a sampling flight	1 s
B	Bandwidth	1 MHz
P	Transmit power	20 dB
E_b	Energy per bit	500 kJ
fUE	Frequency of the user equipment (UE)	0.6 GHz
L, W, H	Length, width, and height of an UAV, respectively	100 *m*
T	Time taken for a single flight	400 s
*V*max	Maximum velocity or speed	50 m/s
a_0	Path loss	− 50 dB
*r*2	Reference received power or signal power	− 100 dBm
c	*constant*	100
S	Sensing range or number of cycles per bit	1000 cycles/bit
fUAV	Frequency of the unmanned aerial vehicle (UAV)	1.2 GHz

2. Train the DQN deep learning model using collected data. The Q-function is initialized to zero. For each episode, the algorithm observes the current state, selects an action using an epsilon-greedy policy, takes action, observes the reward and the next state, and updates the Q-values using the Bellman equation. After the convergence of the DQN model, the optimal action can be selected.
3. Deploy the optimized DQN model in the RIS-assisted UAV communication system to improve the energy efficiency and performance of the communication links.

3.2. Ddpg-based algorithm for continuous cases

In this part, we exhibit a DDPG-based strategy for managing [illegible] happening and upgrading the UAV's direction utilizing the no[illegible]le entertainer pundit approach. Fig. 3 portrays the design of the D[illegible] calculation. There are two DNNs: the entertainer or[illegible]tion, wh[illegible] has the capability a = (s|θ^{μ}), and the pundit organi[illegible]n, [illegible]h has th[illegible] capability Q(s, a|θ^{Q}) individually.

The specialist gets the climate's state s_t an[illegible]veys[illegible] created by its entertainer organization. Th[illegible]ogre[illegible]n saved [illegible]e experience replay memory. *M* change[illegible] examined[illegible]e the growing experience starts to prepare the en[illegible]r and pundi[illegible]anizations. More specifically, the pundit n[illegible]rk c[illegible]s the Q-valu[illegible] $Q(s, a)$ for computing the approach incl[illegible]ation given [illegible]tates' s and activities *a* [19].

4. Result and disc[illegible]n

This section[illegible]es mat[illegible]tical an[illegible]ogrammable [25,26,27] replications to[illegible]ate the[illegible]G[illegible]ed design for process offloading in the U[illegible]assiste[illegible]IEC fram[illegible]k. First, the design of reenactment borde[illegible]provid[illegible]. The performance of the DDPG-based system is then evaluate[illegible]ns and compared to alternative gauge techniques.

4.1. Working setu[illegible]

In this section, we conduct extensive simulations to assess the performance of the proposed algorithms. The simulations are implemented using Python 3.7 and TensorFlow 1.15.0. For the DQN-based algorithm, we employ two fully connected hidden layers, each consisting of 256 neurons. The Adam Optimizer updates the evaluation network with a learning rate of 0.001. The target network is updated with over 300 iterations. In the case of the DDPG-based algorithm, both the actor and critic networks feature two fully connected hidden layers, each comprising 256 neurons. The Adam Optimizer trains both the actor and critic networks with a learning rate of 0.001. The experience replay

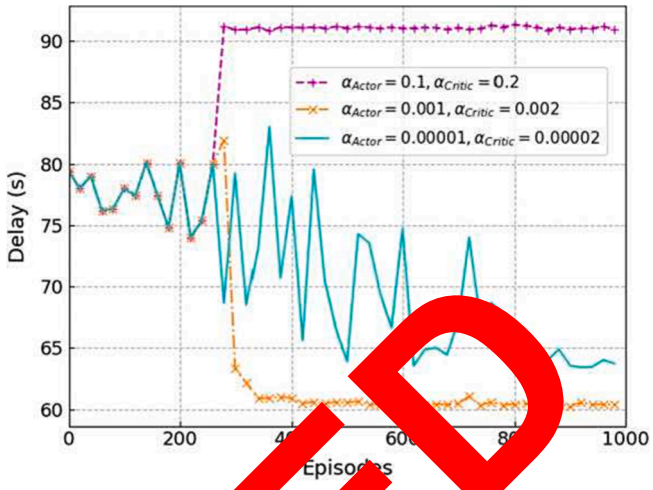

Fig. 4. The union presenta[illegible]f t[illegible]DPG-bas[illegible]d calculation at different learning rates.

memory size i[illegible]100,000, an[illegible]batches of size 128 are used. The coordi[illegible]s of [illegible]ISs are predetermined as [100, −100, 50], [300, 400, 50], [500, −100, 5[illegible] each training episode, the UAV initiates its servi[illegible] User Equipm[illegible]UE) from the starting coordinate [0, 300, 50[illegible]here is a total of 100 T[illegible]nsmitter Stations (TSs) present. The UE, w[illegible]h could be an autonomous vehicle, moves at a constant speed of 6 m[illegible]starting fro[illegible]e initial coordinate [0, 150] and reaching the final co[illegible]nate [600,[illegible]0].

4.2. Simi[illegible]t setting

[illegible]tty gritty reproduction boundaries are recorded in Table 1, except if generally determined. In our trials, the typical award for performance testing, data from many runs of Eq. (2) with identical settings are used. Four standard approaches are depicted as observers for examination:

4.3. Simulation results and discussions

4.3.1. Parametric analysis

We start by progressing preliminaries to decide the ideal credits for significant hyper-limits used in connection estimations. The gathering execution of the proposed computation with fluctuating learning rates is portrayed in Fig. 4. We guess that the pundit organization and entertainer networks have different learning rates. We assume that distinct learning rates are assigned to the critic and actor networks. Firstly, it's evident that the proposed algorithm achieves convergence when Actor = 0.1 and Critic = 0.2 and when Actor = 0.001 and Critic = 0.002. However, when Actor = 0.1 and Critic = 0.2, the algorithm converges to a local optimal solution. This occurs because the large learning rates prompt both the critic and actor networks to take significant update steps simultaneously. Secondly, it's noteworthy that the algorithm fails to converge when the learning rates are exceptionally small, such as Actor = 0.00001 and Critic = 0.00002. This is because lower learning rates result in a slower update rate for the Deep Neural Networks (DNNs), requiring a greater number of iterative episodes to achieve convergence. Consequently, the optimal learning rates for the actor and critic networks are determined to be Actor = 0.001 and Critic = 0.002, respectively.

In Fig. 5, we dissect the impact of various rebate factors c on the proposed calculation's assembly execution. The learned computation offloading policy performs best when the discount factor c 14 0:001 is used. The reason for this is that the environment alters substantially over time. Hence, the information from a solitary time span cannot wholly address long-haul conduct, bringing about unfortunate speculation

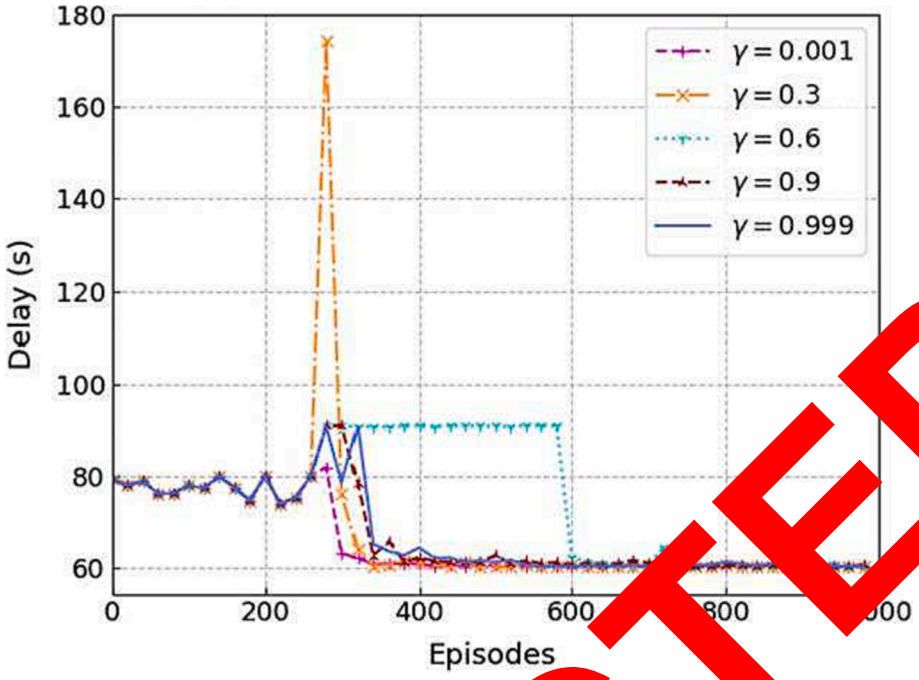

Fig. 5. Performance of convergence w[illegible]arious discount factors.

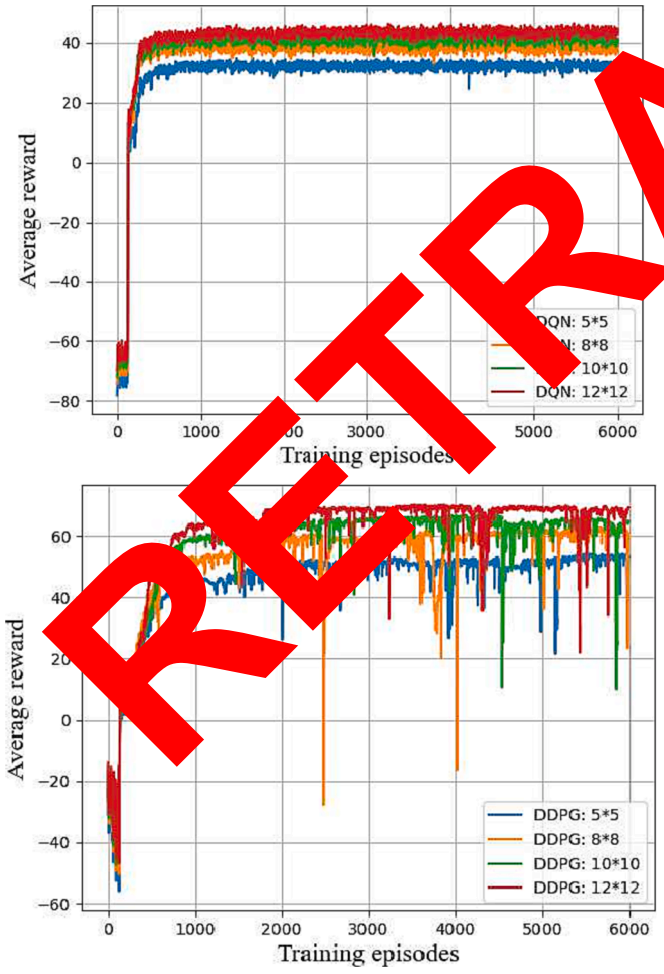

Fig. 6. (a) DQN and (b) DDPG average reward versus the number of training episodes.

c[illegible]city across ti[illegible]spans. Subsequently, a fitting worth of c will expand th[illegible]ssible pre[illegible]tation of our preparation strategies, and in the acco[illegible]s, we set the rebate factor c to 0.001.

As sh[illegible]n Fig. 6(a), the preparation bends of normal prizes are [illegible]rally negative toward the beginning for different quantities of [illegible]components. This is because of the UAV's lackluster showing, for example, flying beyond the objective region, which brings about a negative prize. Following that, when the organizations start to meet, the normal prizes climb and finally settle, showing that the framework has tracked down the best execution. Moreover, one can see that as the quantity of reflecting pieces increments, so do the typical prizes. Then, at that point, in Fig. 6 (b), we show the conventional compensation of the proposed DDPG-based assessment versus how much preparing episodes, which seeks after a near heading as the DQN-based plan in Fig. 6 (a). At the point when the quantity of reflecting components is something very similar, the DDPG-based arrangement gets a preferred compensation over the DQN-based arrangement, true to form. This is on

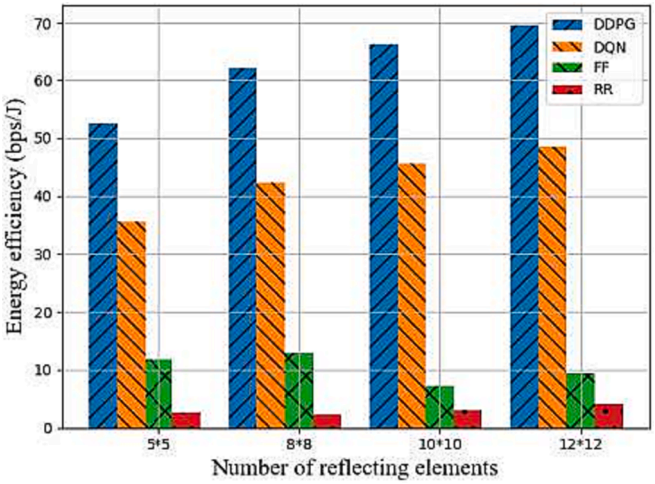

Fig. 7. DQN, DDPG, FF, and RR with varying numbers of reflecting elements attain average energy efficiency.

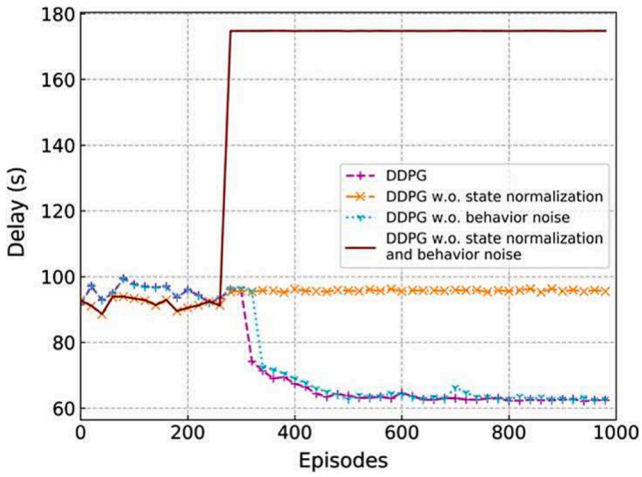

Fig. 8. Performance without state normalization or behavior noise.

the rounds where DQN-based calculations attempt a predetermined number of tasks, while DDPG-based frameworks continually streamline the factors.

4.3.2. Performance comparison

To serve as a benchmark for comparison, we introduce two algorithms:

- Random movement and random phase shifts (RR): This algorithm has the UAV randomly select flying actions and phase shifts for each reflecting element in each time step.
- Fixed movement and fixed phase shifts (FF): This algorithm in… the UAV moving from an initial coordinate of [0, 300, 50] to a… coordinate of [600, 0, 50].

The standard energy efficiency of UAVs obt… ng DQN, DDPG, FF, and RR in a solitary episode is … ayed in …g. 7. For instance, the energy viability of UAVs accomp… by … over the long haul from 52 bps/J to 70 b… . Mo… for fluctuated quantities of reflecting components, D… generally … mplishes better energy productivity when looke… different ca… ions. DQN is somewhat more regrettable th… DDP… hich is additi…ally better compared to FF and RR.

Fig. 8 illustrates how … ing strategies in… the DDPG training algorithm when state n… alizatio… or behavior n…se is not used. If the DDPG algorithm is… ed wi…t behavior noise, its convergence speed will be slower. … y, if the … licy is trained without state normalization … aning … o scale … or is introduced during state normalizati… orithm … unction properly.

Fig. … and 9 … delineate… postponement and offloading proporti… ecution … imilar groups of tests. Fig. 9(a) portrays the union presenta… pire and the DDPG plot with shifting UE registering … cities. The proposed plan is not contrasted with the air conditioner p… cause the air conditioner conspires as yet not concurrent. We can … hat while the registering capacity of the UE is unobtrusive, i.e., f_{UE} 14 0:4 GHz, the handling postpones improved by the two streamlining procedures is more than while the processing ability of the UE is huge, i.e., f_{UE} 14 0:6 GHz. At the point when the UE's computational ability is high, the normal offloading proportion in the framework is low, as displayed in Fig. 9(b), and the UE likes to execute the work locally. The lesser the figuring capacities of the UE, the slower the framework's information handling speed, bringing about a greater, most extreme postponement between neighborhood execution and offloading. Fig. 9(c) portrays an enhanced postpone correlation of the proposed system with the DQN conspire under different central processor recurrence conditions. Fig. 9c shows that the proposed DDPG system accomplishes decreased delay than the DQN strategy under various UE registering capacities. This is because the DDPG scheme can generate a high number of continuous actions, whereas DQN can only generate a small number of discrete actions. Thus, DDPG can discover an accurate component, the offloading ratio, that has a considerable impact on the delay in the continuous action control system.

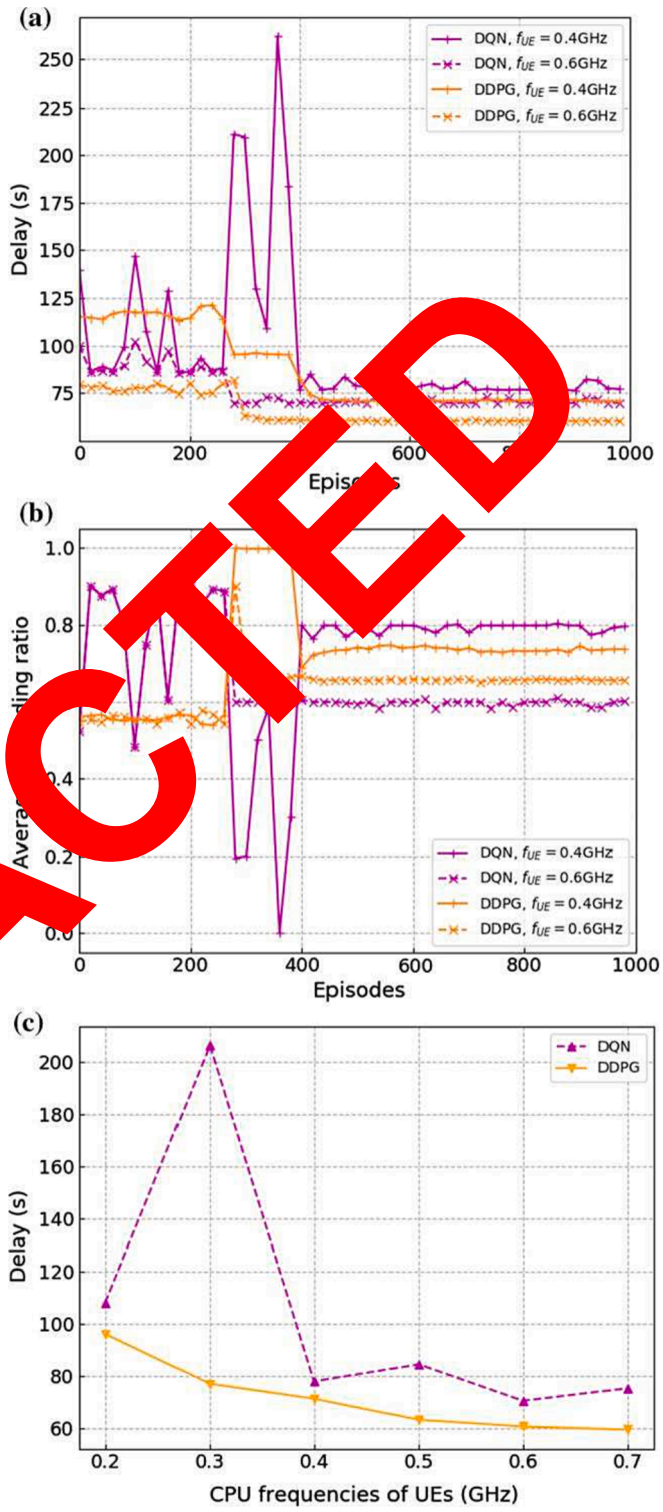

Fig. 9. (a) DQN and DDPG performance under varying UE computational capabilities. (b) DQN and DDPG offloading ratios under varied UE computational capabilities. (c) A comparison of DQN and DDPG.

5. Conclusion

In this paper, we have enhanced the durability of UAV-RISs by employing an energy harvesting technique for SWIPT, resource allocation, and energy extraction from impinging RF signals. Additionally, we have developed the DQN and DDPG techniques to allocate UAV-RIS resources in both the temporal and spatial domains, aiming to maximize the overall harvested energy while maintaining communication quality for each user. The simulation results demonstrate a significant performance improvement of the proposed UAV-RIS SWIPT system compared to the benchmark. In our future research endeavors, we plan to assess the performance of our algorithm in conjunction with other widely recognized RL algorithms, such as actor-critic with experience replay (ACER), actor-critic using Kronecker-factored trust region (ACKTR), proximal policy optimization 2 (PPO2), and trust region policy optimization (TRPO). This exploration aims to comprehensively understand how our approach compares to and complements these established RL techniques.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Aldababsa M, Salhab AM, Nasir AA, Samuh MH, da Costa DB. Multiple RIS networks: Performance analysis and optimization. IEEE Trans Veh Technol 72(6):7545–59. https://doi.org/10.1109/TVT.2023.3242046.

[2] Aldababsa M, Khaleel A, Basar E. STAR-RIS-NOMA networks: An error performance perspective. IEEE Commun Lett 2022;26(8):17 https://doi.org/10.1109/LCOMM.2022.3179731.

[3] Diamanti M, Tsiropoulou EE, Papavassiliou S. IEEE 23 International workshop on signal processing advances in wireless communications (SPAWC Lucca, Italy 2021;2021:621–5. https://doi.org/10.1109/SPAWC

[4] ELMESEIRY, Nourhan, ALSHAER, Nancy, et IS L, Ta tailed survey and future directions of unmanned aerial vehicles (uavs) with ial applications. *Aerospace* 8 12 2021 363.

[5] KURU, Kaya. Planning the future of with swarms of autonomous unmanned aerial vehicles using a ram *IEEE Access* 9 2 6571 6595.

[6] OTTO, Alena, AGATZ, Niels, CAMPBELL, James Optimization approaches for civil applications of unmann rial vehicles (U aerial drones: A survey. *Networks* 72 4 2018 411

[7] Balakrishnan A, De S, L-C. Traffic skewness-aware performance analysis of dual-powered green ar netw EEE Glob. Commun. Conf. Taipei, Taiwan: (GLOBECOM); 202

[8] Peng H, Tsai A-H, Wang L-C, Han Z. LEOPARD: Parallel optimal deep echo state network prediction improves service coverage for UAV-assisted outdoor hotspots. IEEE Trans Cogn Commun Netw 2021. https://doi.org/10.1109/TCCN.2021.3115765.

[9] JI, Jiequ, ZHU, Kun, et CAI, Lin. Trajectory and communication design for cache-enabled UAVs in cellular networks: A deep reinforcement learning approach. *IEEE Transactions on Mobile Computing*, 2022.

[10] Lei M, Zhang X, Yu B, Fowler S, Yu B. Throughput maximization for UAV-assisted wireless powered D2D communication networks with a hybrid time division duplex/frequency division duplex scheme. Wirel Netw 2021;27:2147–57.

[11] ANUMULA, Swarnalatha et GANESAN, Anitha. Wireless power charging of drone using vision-based navigation. *The Journal of Navigation* 74 4 2021 838 852.

[12] Mei H, Yang K, Shen J, Liu Q. Joint trajectory-task-cache optimization with phase-shift design of RIS-assisted UAV for MEC. IEEE Wireless Commun Lett, Apr 2021.

[13] Wu Q, Zhang R. Joint active and passive beamfor ization for intelligent reflecting surface assisted SWIPT under QoS co el Areas Commun 2020;38(8):1735–48.

[14] Samir M, Elhattab M, Assi C, Sharafeddi rayeb A. Opti age of information through aerial reconfigurable nt surfaces: A reinforcement learning approach. IE Trans chnol 2021):3978–83.

[15] Ye H, Li GY, Juang B-H-F. Deep cement le based rce allocation for V2V communications. IEE s Veh Technol 2 63–73.

[16] SHANG, Bodong, SHAFIN, et et LI gjia. U n-enabled aerial reconfigurable intelligent e (SAR *EE Wireless Communications* 28 5 2021 156 163.

[17] Wang H, Wang J, D Wan sis TA, Sh K. "Resource allocation for energy harvesti ered D2D nicati derlaying UAV-assisted networks. IE s Green Commu ;2(1):14–24.

[18] Y. Tang G L. Xu X. Han "Jo smit and reflective beamforming design fo -assis ltiuser MISO SW T systems" in IEEE Int. Conf. Commun. (ICC), Dublin, Ireland 2020.

[19] A. Al-Hourani S. Kandeep Jamalipour "Modeling air-toground path loss for de platforms in u vironments" in IEEE Glob. Commun. Conf. ustin, TX, Dec. 2014.

[20] A. Al-Hourani, S. Kandeepan, S. Lardner. "Optimal LAP altitude for maximum coverage" IEE eless Commun. Lett., 3 6 569, Dec. 2014.

[21] DIAMANTI, Ma TSAMPAZI, Maria, TSIROPOULOU, Eirini Eleni, *et al.* Energy ficient multi- communications aided by reconfigurable intelligent surfaces UAVs. I *21 IEEE International Conference on Smart Computing* (S IEEE, 2021 371 376.

[22] Li S, Đurđevic P, Yang Z. Optimal tracking control based on integral reinforcement learning for an underactuated drone. Elsevier IFAC-PapersOnLine 2019;52(8):

[23] Attwood C, Richards AG. Reinforcement learning and model predictive control for robust embedded quadrotor guidance and control. Springer Autonomous Robots 2019;43(7):1681–93.

[24] Koch W, Mancuso R, West R, Bestavros A. Reinforcement learning for UAV attitude control. ACM Transactions on Cyber-Physical Systems 2019;3(2):1–21.

[25] M. Abadi A. Agarwal P. Barham E. Brevdo Z. Chen C. Citro G. S. Corrado A. Davis J. Dean M. Devin et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems" arXiv preprint arXiv:1603.04467, 2016.

[26] P. Santi, R. Kesav, S. Palaniappan, P. Balaji, D. Loreto, C. Ottavio, S. Radha, N. Prabagarane, S. Gino. Comparative analysis of machine learning and physics-based optimizations of a dual circularly polarized antenna for V2X applications, AEU - International Journal of Electronics and Communications, Dec. 2021.

[27] Kanhaiya S, Ganga P. Efficient modelling of compact microstrip antenna using machine learning. AEU - International Journal of Electronics and Communications 2021.