CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

# Conference Paper

## Deep Reinforcement Learning for Real-Time Trajectory Planning in UAV Networks

**Kai LiWei NiEduardo TovarMohsen Guizani**

CISTER-TR-200401

# Deep Reinforcement Learning for Real-Time Trajectory Planning in UAV Networks

Kai LiWei NiEduardo TovarMohsen Guizani

CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail:

https://www.cister-labs.pt

## Abstract

In Unmanned Aerial Vehicle (UAV)-enabled wirelesspowered sensor networks, a UAV can be employed tocharge the ground sensors remotely via Wireless Power Transfer(WPT) and collect the sensory data. This paper focuseson trajectory planning of the UAV for aerial data collectionand WPT to minimize buffer overflow at the ground sensorsand unsuccessful transmission due to lossy airborne channels.Consider network states of battery levels and buffer lengthsof the ground sensors, channel conditions, and location of theUAV. A flight trajectory planning optimization is formulatedas a Partial Observable Markov Decision Process (POMDP),where the UAV has partial observation of the network states.In practice, the UAV-enabled sensor network contains a largenumber of network states and actions in POMDP while theup-to-date knowledge of the network states is not availableat the UAV. To address these issues, we propose an onboarddeep reinforcement learning algorithm to optimize the realtimetrajectory planning of the UAV given outdated knowledgeon the network states.

# Deep Reinforcement Learning for Real-Time Trajectory Planning in UAV Networks

Kai Li
*CISTER Research Centre*
Porto, Portugal
kai@isep.ipp.pt

Wei Ni
*CSIRO*
Sydney, Australia
wei.ni@data61.csiro.au

Eduardo Tovar
*CISTER Research Centre*
Porto, Portugal
emt@isep.ipp.pt

Mohsen Guizani
*Qatar University*
Doha, Qatar
mguizani@ieee.org

***Abstract*—In Unmanned Aerial Vehicle (UAV)-enabled wireless powered sensor networks, a UAV can be employed to charge the ground sensors remotely via Wireless Power Transfer (WPT) and collect the sensory data. This paper focuses on trajectory planning of the UAV for aerial data collection and WPT to minimize buffer overflow at the ground sensors and unsuccessful transmission due to lossy airborne channels. Consider network states of battery levels and buffer lengths of the ground sensors, channel conditions, and location of the UAV. A flight trajectory planning optimization is formulated as a Partial Observable Markov Decision Process (POMDP), where the UAV has partial observation of the network states. In practice, the UAV-enabled sensor network contains a large number of network states and actions in POMDP while the up-to-date knowledge of the network states is not available at the UAV. To address these issues, we propose an onboard deep reinforcement learning algorithm to optimize the real-time trajectory planning of the UAV given outdated knowledge on the network states.**

***Index Terms*—Wireless sensor networks, Unmanned aerial vehicles, Trajectory planning, Wireless power transfer, Deep reinforcement learning**

## I. Introduction

In recent years, Unmanned Aerial Vehicles (UAVs), also known as drones, are employed to collect and process data from ground sensors deployed for environmental monitoring, e.g., surveillance of land pollution or air quality in smart cities [1], [2]. The ground sensors, equipped with a data communication antenna and a wireless power receiver for Wireless Power Transfer (WPT) [3], generate data packets at an application-specific sampling rate, and put them into a data buffer for future transmission. Thanks to the excellent maneuverability, the UAV can move sufficiently close to each ground sensor, exploiting short-distance line-of-sight (LoS) communications for data collection [4] and batteries charging via WPT [5]. Moreover, the UAV with a large size can be equipped with lightweight high-capacity rechargeable batteries and solar panels on top of its wings [6]. By harvesting solar energy to charge the batteries, the UAV can fly over a long distance without landing and hover in the air for an extended period.

Figure 1 presents a typical UAV-enabled sensor network, where a number of sensor nodes serve as data sources with sensing ability for monitoring the urban environment in smart cities. The UAV equipped with a high-capacity battery, wireless radio, onboard processors, and a WPT transmitter hovers over the area of interest. The UAV adapts its flight trajectory for collecting or ferrying the data of the ground sensors given limited radio coverage of the UAV and the ground sensors. Battery energy of the ground sensors can be greatly different from each other, since the amount of harvested energy via WPT can be impacted by natural conditions, e.g., weather or wireless interference from existing wireless networks. Moving to the ground sensor with poor channel condition or small buffer length for data collection and WPT gives rise to packet reception errors or buffer overflow at other ground sensors. Moreover, in practice, battery energy levels, data buffer lengths, and channel conditions of all the ground sensors are not available or can only be partially observed by the UAV. Therefore, online trajectory planning of the UAV for data collection and WPT, for preventing data lost resulting from buffer overflow and fading channels is crucial in UAV-enabled sensor networks.

In this paper, we first formulate the real-time trajectory planning of the UAV for aerial data collection and WPT as a Partial Observable Markov Decision Process (POMDP). Each POMDP state contains battery levels and data buffer lengths of the ground sensor, channel conditions, and waypoints of the UAV along the trajectory. Then, a new Deep Reinforcement Learning based Trajectory Planning (DRLTP) algorithm is developed, which derives the optimal instantaneous waypoints of the UAV according to the network states, actions and a corresponding Q value. DRLTP learns the optimal Q value asymptotically through training a deep Q-network onboard the UAV to optimize the flight trajectory.

This paper is structured as follows. Related work on trajectory planning of the UAV is presented in Section II. Formulation of air-ground channel, flight trajectory, and energy model is investigated in Section III. The onboard deep reinforcement learning algorithm for trajectory planning is proposed in Section IV to minimize the packet loss. In Section V, we show numerical results and performance evaluation. This paper is concluded in Section VI.

## II. Related Work

This section presents the literature on trajectory planning of the UAV and WPT. The trajectory of the UAV is designed to increase spectral efficiency of small cells in cellular networks [7]. The UAV can make movement decisions using local position information of the active user that it is currently serving, or requiring information about the user locations in neighbor cells. In [8], a UAV-assisted mobile edge computing (MEC) network is studied, in which the UAV moving around the target serves as a computing server to process the ground users' tasks or relay their requests. A resource allocation algorithm is presented to reduce the energy consumption of the UAV and the ground users, subject to the computation resource scheduling, bandwidth allocation, and the UAV's trajectory. In bandwidth hungry and delay-tolerant sensor networks, a communication framework is developed in [9] to assign UAVs to assist data transmission of the ground sensors. The trajectory of the UAV is determined such that the total service time can be reduced. In [10], the trajectory design of the UAV is formulated as a mixed integer non-linear programming problem to reduce the average user-to-UAV pathloss, considering the state-of-the-art air-ground channel model. A trajectory planning algorithm is developed, in which the sub-problems are solved iteratively through the block coordinate descent method.

A two-user UAV-enabled WPT system is studied in [11], where the trajectory of the UAV is designed to characterize the Pareto boundary of the achievable energy region of the energy receivers. Based on the distance between the UAV and the energy receivers, the boundary of the energy region can be determined by either hovering above a fixed location or a hover-fly-hover trajectory. Under the maximum flying speed constraint, designing the trajectory of the UAV can also enhance the minimum harvested energy among all ground nodes in the UAV-enabled WPT system [12]. In [13], a UAV-enabled wireless powered MEC system is presented, where the UAV provides multiple ground users with computation offloading and WPT services. The energy consumption of the UAV is reduced by adjusting the offloading computation bits, the CPU frequency of users, and the trajectory of the UAV. Since the computation performance and the harvested energy can be impacted by propagation loss, the UAV trajectory is also designed to improve computation rates of users in the UAV-enabled MEC wireless powered system [14]. In [15], deep reinforcement learning is studied for aerial data collection in the sensor networks, where the circular trajectory of the UAV is predetermined and maintained. The UAV equipped with a WPT transmitter schedules the ground sensor within the radio coverage to transmit data for preventing buffer overflows while charging the battery of ground sensor to extend network lifetime.

The existing trajectory planning and WPT approaches in the UAV networks focus on the formulation with fully observable network states. However, the practical scenario with the partial observable states of the energy harvesting and the data collection was yet to be considered.

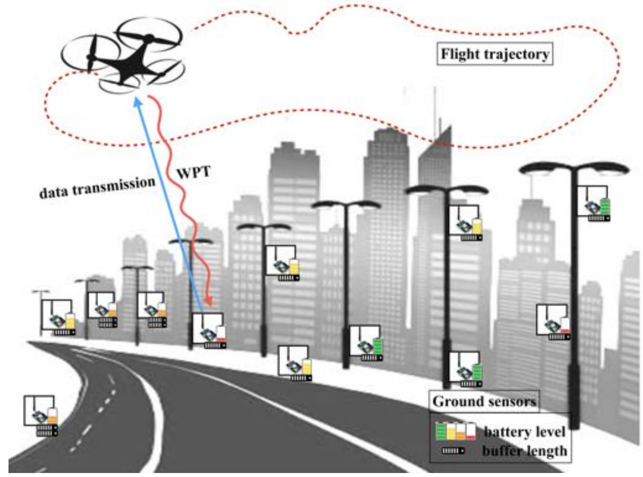


Fig. 1: Employing a UAV for aerial data collection and WPT, where wireless powered ground sensors are deployed for environmental monitoring in the smart city.

## III. Network Model

In this section, we formulate the UAV-ground channel, flight trajectory, and WPT model of the ground sensor.

### A. Channel model

The network under consideration consists of $N$ geographically distributed wireless powered sensor nodes. The UAV that acts as a data collection node flies a predetermined trajectory for $L$ laps. The flight waypoint of the UAV at $t$ in lap $l$ is denoted by $w_l(t)$. The UAV uses WPT to remotely charge the ground sensors. The sensor $i$ ($\in [1, N]$) harvests energy from the UAV to power its operations, e.g., sensing, computing and communication. The rechargeable battery of the sensor is finite with the capacity of $E$ Joules, and the battery overflows if overcharged. In particular, $\rho_i^l(t)$ = 1, 2, and 3 indicates binary phase-shift keying (BPSK), quadrature-phase shift keying (QPSK), and 8 phase-shift keying (8PSK), respectively, and $\rho_i^l(t) \geq 4$ provides $2^{\rho_i^l(t)}$ quadrature amplitude modulation (QAM).

According to the Nakagami-$m$ channel model [16], the required BER of ground sensor's data transmission, denoted by $\varepsilon$, can be given by

$$\varepsilon \approx \frac{0.2}{\Gamma(m)}(\frac{m}{\bar{\gamma}_i})^m \Big[\frac{\Gamma(m, b_{\rho_i^l(t)}\gamma_i(\rho_i^l(t)))}{(b_{\rho_i^l(t)})^m} - \frac{\Gamma(m, b_{\rho_i^l(t)}\gamma_i(\rho_i^l(t)+1))}{(b_{\rho_i^l(t)})^m}\Big], \quad (1)$$

$$b_{\rho_i^l(t)} = \frac{m}{\bar{\gamma}_i} + \frac{3}{2(2^{\rho_i^l(t)}-1)}, \quad (2)$$

where $\Gamma(\cdot)$ is the Gamma function [17], $\gamma_i(\rho_i^l(t))$ is the SNR between node $i$ and the UAV using $\rho_i^l(t)$, and the average SNR of the channel is $\bar{\gamma}_i$ [18]. $\rho_i^l(t)$ denotes the modulation scheme of node $i$ at $t$ in lap $l$. Moreover, we have $\gamma_i(\rho_i^l(t)) = \frac{\|\mathbf{h}_i^l(t)\|^2 P_i^l(t)}{\sigma_0^2}$, where $\sigma_0^2$ is noise power of the channel, and $P_i^l(t)$ is the transmit power of $i$ in lap $l$ [19]. $\mathbf{h}_i^l(t)$ is the complex channel coefficient, which can be known by channel reciprocity.

For illustration convenience, we consider $m = 1$ in (1) as an example in this paper [20]. Note that other Nakagami fading channel model with any $m$ values can also be applied to the proposed DRLTP scheme. Given a specific $\varepsilon$, the transmit power of the ground sensor can be given by [21]

$$P_i^l(t) \approx \frac{\kappa_2^{-1} \ln \frac{\kappa_1}{\varepsilon}}{\|\mathbf{h}_i^l(t)\|^2}(2^{\rho_i^l(t)} - 1), \tag{3}$$

where $\kappa_1$ and $\kappa_2$ denote the two channel constants.

### B. Flight and WPT model

The distance between the UAV and node $i$ along the flight trajectory at $t$ in lap $l$ is $d_i^l(t)$. The WPT transceiver alignment between the UAV and the ground sensor is $\theta_i^l(t)$. Based on [22], we have the WPT efficiency factor $\phi(d_i^l(t), \theta_i^l(t))$, which depends on the distance and the WPT transceiver alignment. The power transferred from the UAV to the ground sensor via WPT can be given by

$$\tilde{P}_i^l(t) = \phi(d_i^l(t), \theta_i^l(t)) P_{\mathrm{UAV}}^{tx} \|\mathbf{h}_i^l(t)\|^2, \tag{4}$$

where the transmit power of WPT at the UAV is $P_{\mathrm{UAV}}^{tx}$, and $\|\cdot\|$ denotes norm.

Each communication frame consists of a number of equally divided time slots, in which the UAV moves to one new waypoint while a sensor node transmits a data packet to the UAV in each slot. The proposed DRLTP is carried out on the UAV to determine instantaneously the next waypoint along the trajectory, where the ground sensor with the best channel harvests energy and transmits its sensory data. Then, the UAV transfers power to the ground sensor via WPT, followed by the data transmission from the ground sensor to the UAV. In particular, a control segment of the ground sensor's data packet contains its battery level and buffer length. Additionally, the updated state information of the ground sensor can be known by the UAV based on the received data.

## IV. Deep Reinforcement Learning for Trajectory Planning

Consider a practical scenario, where the UAV has only partial observation of the network states. The trajectory planning of the UAV for the data collection with WPT is formulated as a POMDP. Given a large action space in POMDP, DRLTP is proposed to minimize packet loss of the ground sensors, by optimizing the flight trajectory.

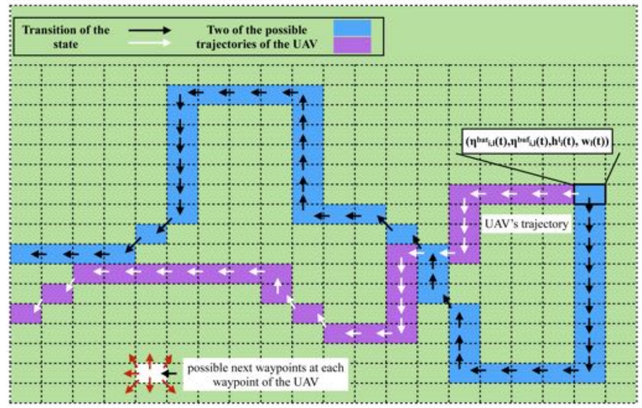


Fig. 2: Each block denotes the possible waypoint of the UAV. At each waypoint, the proposed DRLTP scheme determines the next waypoint of the UAV to minimize the packet loss of the ground sensors.

### A. POMDP formulation

Each POMDP state $\alpha$ is composed of battery levels $\eta_{i,l}^{\mathrm{bat}}(t)$ and buffer lengths $\eta_{i,l}^{\mathrm{buf}}(t)$ of the ground sensor, channel conditions $\mathbf{h}_i^l(t)$, and waypoints $w_l(t)$ of the UAV along the flight trajectory [23]. Thus, we have $\alpha = \{\eta_{i,l}^{\mathrm{bat}}(t), \eta_{i,l}^{\mathrm{buf}}(t), \mathbf{h}_i^l(t), w_l(t)\}$. In the proposed POMDP formulation, the action $\mathcal{A}_\alpha$ to be taken is to determine $w_l(\beta)$, where the ground sensor with the best channel condition harvests energy via WPT and transmits data to the UAV.

In practical scenarios, the UAV has no full observation of $\eta_{i,l}^{\mathrm{bat}}(\alpha)$, $\eta_{i,l}^{\mathrm{buf}}(\alpha)$ and $\mathbf{h}_i^l(\alpha)$ at state $\alpha$. The measurements of the network state can be corrupted by random errors, such as $\eta_{i,l}^{\mathrm{bat}\prime}$, $\eta_{i,l}^{\mathrm{buf}\prime}$, and $\mathbf{h}_i^{l\prime}$, that are spatially varying, i.e., the measurement error covariance depends on the movements of the UAV. Thus, we define $\alpha_l'$ as a partial observation of the network states, where

$$\alpha_l' = \{\eta_{i,l}^{\mathrm{bat}}(\alpha) + \eta_{i,l}^{\mathrm{bat}\prime}, \eta_{i,l}^{\mathrm{buf}}(\alpha) + \eta_{i,l}^{\mathrm{buf}\prime}, \mathbf{h}_i^l(\alpha) + \mathbf{h}_i^{l\prime}\} \tag{5}$$

A feature of POMDPs is that the state of the process is not represented by a single value but by a probability distribution over all the possible states in its state-space representation at a particular time, this is known as a belief-state and is denoted by $\psi$. Moreover, the objective of POMDP is to choose optimal action policy $\pi^*$ over a time horizon, such that the expected cumulative network cost is minimized. Let $\mathcal{C}_\alpha$ denote the network cost, i.e., the packet loss due to the buffer overflows and channel fading, when action $\mathcal{A}_\alpha$ is taken by the UAV. The expected cumulative cost can be written as $\sum_{\alpha \in \mathcal{S}} \lambda^\alpha \mathcal{C}_\alpha$, which is the sum of the discounted network cost after executing every action. $\lambda \in [0, 1]$ is a discount factor for future states, and $\mathcal{S}$ gives the state space. The expected return from the belief state $\psi$

according to policy $\pi$ in the formulated POMDP defines

$$\Omega^{\pi}(\psi) = \min_{\pi \in \Pi} \ \mathbb{E}^{\pi}\left\{ \sum_{\alpha \in \mathcal{S}} \lambda^{\alpha} \mathcal{C}_{\alpha} \ \middle| \ \mathcal{A}_{\alpha}, \psi \right\} \tag{6}$$

Therefore, the optimal policy for POMDP is the one which minimizes the objective, which gives

$$\pi^{*}(\psi) = \arg\min_{\pi} \Omega^{\pi}(\psi) \tag{7}$$

Figure 2 presents transition diagram of POMDP states and actions with two possible trajectories as an example, in which the blocks stand for all possible states in $\mathcal{S}$ in POMDP, i.e., the waypoints $w_l(t)$ in the space. At each waypoint, where $\alpha = \{\eta_{i,l}^{\text{bat}}(\alpha), \eta_{i,l}^{\text{buf}}(\alpha), \mathbf{h}_i^l(\alpha), w_l(\alpha)\}$, the UAV can move along one of the 7 possible directions to the next state/waypoint, i.e., $\beta$. According to the state transition probability, the state transition depends on the change of $\{\eta_{i,l}^{\text{bat}}(\alpha), \eta_{i,l}^{\text{buf}}(\alpha), \mathbf{h}_i^l(\alpha)\}$ of the ground sensor and $w_l(\alpha)$ on the trajectory of the UAV. For example, consider a small example of 16 MDP states in one lap of the UAV's flight, where we have 3 waypoints on the trajectory, 2 ground sensors, and the channel is constant. Each sensor has 2 battery levels, and data buffer can hold 2 packets. The next state of $\alpha = \{\eta_{i,1}^{\text{bat}}(\alpha) = 1, \eta_{i,1}^{\text{buf}}(\alpha) = 1, w_l(\alpha) = (x_1, y_1)\}$ can be $\{\eta_{i,1}^{\text{bat}}(\beta) = 1, \eta_{i,1}^{\text{buf}}(\beta) = 2, w_l(\beta) = (x_2, y_2)\}$, when the UAV moves to the next waypoint $(x_2, y_2)$ with two possible states of the ground sensor which does not successfully harvest energy, i.e., (i) node $i$ is selected, but the data transmission is not successful; (ii) node $i$ is not selected and a new packet arrives at node $i$. $\beta$ can also be $\{\eta_{i,1}^{\text{bat}}(\beta) = 2, \eta_{i,1}^{\text{buf}}(\beta) = 1, w_l(\beta) = (x_3, y_3)\}$, when the UAV moves to $(x_3, y_3)$ with two possible states of node $i$ which successfully harvest energy, i.e., (i) node $i$ is selected, and the data transmission is successful; (ii) node $i$ is not selected and there is no new packet arrival. Note that Figure 2 gives a small-scale example of the state transition with two possible trajectories. The UAV can have thousands of possible trajectories and the problem can contain over thousands of POMDP states, which leads to an extremely complex state transition diagram.

### *B. DRLTP algorithm*

A deep Q-network is designed by the proposed DRLTP onboard the UAV to optimize $w_l(t)$ on the flight trajectory of the UAV by approximating the optimal Q value for the data collection with WPT. Figure 3 depicts the proposed deep Q-network, where the Q value is derived according to $\alpha = \{\eta_{i,l}^{\text{bat}}(t), \eta_{i,l}^{\text{buf}}(t), \mathbf{h}_i^l(t), w_l(t)\}$ and the actions of the UAV. The learning weight values $\varphi_j$ in the deep Q-network are iteratively adjusted to approximate $Q\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}; \varphi_j\right\}$, where $j \leq J$ and $J$ is the total number of iterations. The approximated $Q\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}; \varphi_j\right\}$, which is the outputs of the deep Q-network, can be minimized by optimizing the weights $\varphi_j$.

The weight $\varphi_j$ at iteration $j$ is adjusted for training the deep Q-network, while minimizing $Q\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}; \varphi_j\right\}$. At

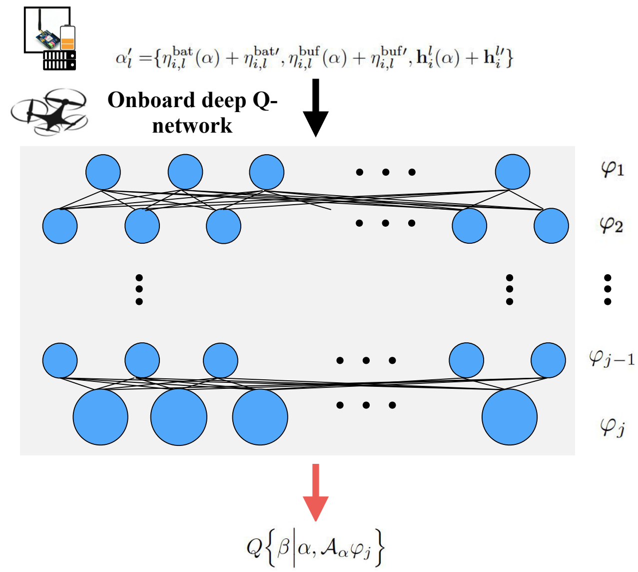


Fig. 3: The deep Q-network structure in DRLTP onboard the UAV.

each iteration of minimizing $Q\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}; \varphi_j\right\}$, the weight $\varphi_{j-1}$ from iteration $(j-1)$ is fixed. Thus, the subproblem of learning $Q\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}; \varphi_j\right\}$ at iteration $j$ $(j \leq J)$ defines $C\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}\right\} + \delta \min_{\mathcal{A}_{\alpha}' \in \mathcal{A}} Q\left\{\beta' \middle| \beta, \mathcal{A}_{\alpha}'; \varphi_{j-1}\right\}$. For deriving the weight $\varphi_j$ at iteration $j$, gradient descent is applied to iteratively compute the gradient value, and update the neural network's weights to reach the global minimum (refer to [15] for details).

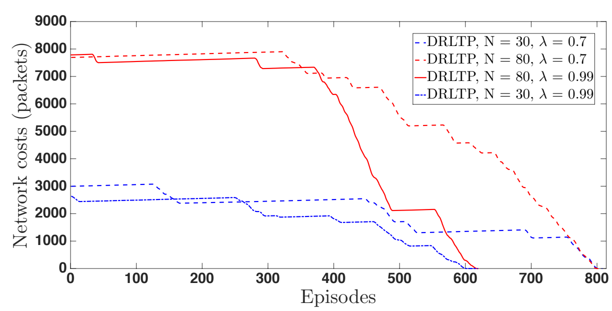


Fig. 4: Data packet loss with regards to training DRLTP, where $\lambda = 0.99$ or $0.7$.

The proposed DRLTP scheme optimizes the actions of the UAV based on the deep Q-network to solve the optimization problem of trajectory planning for minimizing the data loss of the wireless powered ground sensors. The proposed deep Q-network maintains two separate Q-networks $Q\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}\varphi_j\right\}$ and $Q\left\{\beta \middle| \alpha, \mathcal{A}_{\alpha}\varphi_{j-1}\right\}$ with the weights $\varphi_j$ at iteration $j$ and the weights $\varphi_{j-1}$ at iteration $j-1$, respectively. DRLTP updates $\varphi_j$ with multiple times per time-step, and $\varphi_j$ is copied into $\varphi_{j-1}$. DRLTP trains the deep Q-network to minimize a set of loss functions at every update iteration, hence, minimizing the mean-squared

Bellman error. Therefore, the optimality can be asymptotically achieved by DRLTP. Note that DRLTP is generic and can work with the advanced flight models [5] and different types of the UAV, e.g., the Bluetooth-connected UAV in [24].

## V. Numerical Results

DRLTP is implemented in Python 3.5 based on Google TensorFlow, and we assess the performance when the number of episodes is 800, and the total number of waypoints enlarges from 10 to 80.

Figures 4 shows the network cost, i.e., packet loss of the ground sensors, where each ground sensor generates 100 data packets and $\lambda$ is set to 0.99 and 0.7, respectively. DRLTP has a high network packet loss at the first 300 episodes during the training given $N = 30$ or 80. Then, the overall packet loss substantially drops as the trajectory planning strategy is optimally trained. In particular, the packet loss of DRLTP with $\lambda = 0.99$ falls to 0 in 638 episodes. This is achieved by training the onboard DQN for minimizing the packet loss. However, the convergence of DRLTP with $\lambda = 0.7$ requires more than 188 episodes. This is because the duration of training the onboard DQN can be reduced by a high discount factor.

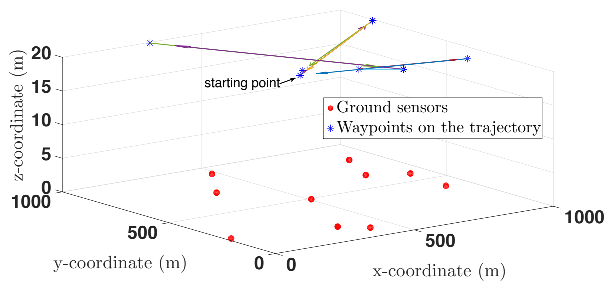


Fig. 5: Training the trajectory planning of the UAV given 10 waypoints.

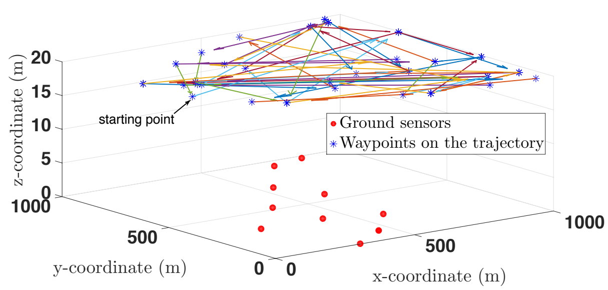


Fig. 6: Training the trajectory planning of the UAV given 50 waypoints.

Figures 5, 6, and 7 show the trajectories at the beginning of training given $J = 10$. DRLTP determines the number of waypoints of the UAV and calculates the Q-function based on the observed states on the UAV (i.e., waypoints, channel conditions) and the ground sensors (i.e., data buffers and battery levels).

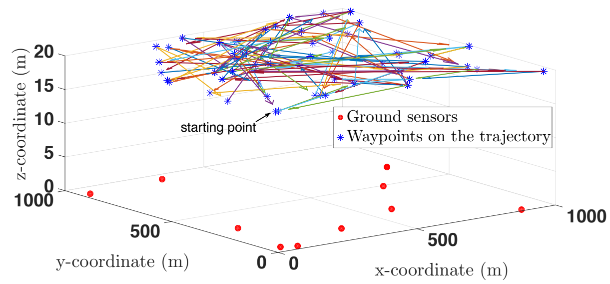


Fig. 7: Training the trajectory planning of the UAV given 80 waypoints.

## VI. Conclusion and Future Work

This paper studies the optimal flight trajectory planning of the UAV for aerial data collection and WPT in sensor networks. The trajectory planning of the UAV is formulated as a POMDP to minimize the data lost due to buffer overflows at the ground sensors and fading airborne channels. Given the large state and action spaces in POMDP, an onboard deep reinforcement learning based trajectory planning algorithm, DRLTP, is proposed to optimally determine the instantaneous waypoints of the UAV for the data collection and WPT. The proposed DRLTP is implemented on Google TensorFlow. Numerical results demonstrate that the trajectories of the UAV are trained by DRLTP, which significantly reduces the data packet loss given different number of ground sensors.

For future work, the ground sensors with dynamic battery capacity and data buffer size will be considered to harvest energy from multiple resources. The proposed DRLTP will be further evaluated in multiple application scenarios, e.g., 5G or vehicular networks.


## Acknowledgements

This work was partially supported by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology), within the CISTER Research Unit (UIDB/04234/2020); also by the Operational Competitiveness Programme and Internationalization (COMPETE 2020) under the PT2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by national funds through the FCT, within project(s) POCI-01-0145-FEDER-029074 (ARNET).


## References


[1] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges," *IEEE Access*, vol. 7, pp. 48 572–48 634, 2019.

[2] O. A. Alvear, N. R. Zema, E. Natalizio, and C. T. Calafate, "A chemotactic pollution-homing UAV guidance system," in *International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, 2017, pp. 2115–2120.

[3] K. Li, W. Ni, M. Abolhasan, and E. Tovar, "Reinforcement learning for scheduling wireless powered sensor communications," *IEEE Transactions on Green Communications and Networking*, 2018.
[4] B. Li, Z. Fei, and Y. Zhang, "UAV communications for 5G and beyond: Recent advances and future trends," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241–2263, 2018.
[5] K. Li, W. Ni, E. Tovar, and A. Jamalipour, "Deep Q-learning based resource management in UAV-assisted wireless powered IoT networks," in *International Conference on Communications (ICC)*. IEEE, 2020.
[6] Y. Emami, K. Li, and E. Tovar, "Buffer-aware scheduling for uav relay networks with energy fairness," in *Vehicular Technology Conference (VTC Spring)*. IEEE, 2020.
[7] A. Fotouhi, M. Ding, and M. Hassan, "Dynamic base station repositioning to improve spectral efficiency of drone small cells," in *IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2017, pp. 1–9.
[8] X. Hu, K.-K. Wong, K. Yang, and Z. Zheng, "UAV-assisted relaying and edge computing: Scheduling and trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4738–4752, 2019.
[9] H. Ghazzai, M. B. Ghorbel, A. Kassler, and M. J. Hossain, "Trajectory optimization for cooperative dual-band UAV swarms," in *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.
[10] W. Shi, J. Li, N. Cheng, F. Lyu, Y. Dai, H. Zhou, and X. S. Shen, "3D multi-drone-cell trajectory design for efficient IoT data collection," in *IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.
[11] J. Xu, Y. Zeng, and R. Zhang, "UAV-enabled wireless power transfer: Trajectory design and energy region characterization," in *IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2017, pp. 1–7.
[12] Y. Hu, X. Yuan, J. Xu, and A. Schmeink, "Optimal 1D trajectory design for UAV-enabled multiuser wireless power transfer," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5674–5688, 2019.
[13] F. Zhou, Y. Wu, H. Sun, and Z. Chu, "UAV-enabled mobile edge computing: Offloading optimization and trajectory design," in *IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.
[14] F. Zhou, Y. Wu, R. Q. Hu, and Y. Qian, "Computation rate maximization in UAV-enabled wireless-powered mobile-edge computing systems," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 9, pp. 1927–1941, 2018.
[15] K. Li, W. Ni, E. Tovar, and A. Jamalipour, "On-board deep Q-network for UAV-assisted online power transfer and data collection," *IEEE Transactions on Vehicular Technology*, 2019.
[16] M.-S. Alouini and A. J. Goldsmith, "Adaptive modulation over nakagami fading channels," *Wireless Personal Communications*, vol. 13, no. 1-2, pp. 119–143, 2000.
[17] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic press, 2014.
[18] K. Li, R. C. Voicu, S. S. Kanhere, W. Ni, and E. Tovar, "Energy efficient legitimate wireless surveillance of UAV communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2283–2293, 2019.
[19] K. Li, W. Ni, E. Tovar, and M. Guizani, "Optimal rate-adaptive data dissemination in vehicular platoons," *IEEE Transactions on Intelligent Transportation Systems*, 2019.
[20] K. Li, N. Ahmed, S. S. Kanhere, and S. Jha, "Reliable transmissions in AWSNs by using O-BESPAR hybrid antenna," *Pervasive and Mobile Computing*, vol. 30, pp. 151–165, 2016.
[21] K. Li, W. Ni, X. Wang, R. P. Liu, S. S. Kanhere, and S. Jha, "Energy-efficient cooperative relaying for unmanned aerial vehicles," *IEEE Transactions on Mobile Computing*, no. 6, pp. 1377–1386, 2016.
[22] K. Li, C. Yuen, and S. Jha, "Fair scheduling for energy harvesting WSN in smart city," in *SenSys*. ACM, 2015, pp. 419–420.
[23] K. Li, S. S. Kanhere, W. Ni, E. Tovar, and M. Guizani, "Proactive eavesdropping via jamming for trajectory tracking of UAVs," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2019, pp. 477–482.
[24] K. Li, N. Lu, P. Zhang, W. Ni, and E. Tovar, "Multi-drone assisted internet of things testbed based on bluetooth 5 communications," in *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. ACM, 2020.