Trajectory Optimization of Flying Energy Sources using Q-Learning to Recharge Hotspot UAVs

Sayed Amir Hoseini[1], Jahan Hassan[2], Ayub Bokani[3], Salil S. Kanhere[4]
[1,2,3] School of Engineering and Technology, The Central Queensland University, Sydney, Australia
{s.hoseini, j.hassan, a.bokani}@cqu.edu.au
[4] School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia
salil.kanhere@unsw.edu.au

***Abstract*—Despite the increasing popularity of commercial usage of UAVs or drone-delivered services, their dependence on the limited-capacity on-board batteries hinders their flight-time and mission continuity. As such, developing in-situ power transfer solutions for topping-up UAV batteries have the potential to extend their mission duration. In this paper, we study a scenario where UAVs are deployed as base stations (UAV-BS) providing wireless Hotspot services to the ground nodes, while harvesting wireless energy from flying energy sources. These energy sources are specialized UAVs (Charger or transmitter UAVs, tUAVs), equipped with wireless power transmitting devices such as RF antennae. tUAVs have the flexibility to adjust their flight path to maximize energy transfer. With the increasing number of UAV-BSs and environmental complexity, it is necessary to develop an intelligent trajectory selection procedure for tUAVs so as to optimize the energy transfer gain. In this paper, we model the trajectory optimization of tUAVs as a Markov Decision Process (MDP) problem and solve it using Q-Learning algorithm. Simulation results confirm that the Q-Learning based optimized trajectory of the tUAVs outperforms two benchmark strategies, namely random path planning and static hovering of the tUAVs.**

I. Introduction

The rapid advancement and falling costs of Unmanned Aerial Vehicle (UAVs), a.k.a. drones, are fueling their popularity in a wide range of commercial applications such as goods delivery, medical logistics, entertainment, and aerial imagery [1]. The global market of emerging drone-aided commercial services is estimated at a staggering value of $\$127bn$ [2]. Interestingly, the next generation, i.e., 5G and beyond, mobile communication systems are also expected to use UAVs as either aerial base stations/relays/hotspots (UAV-BSs) [3], [4], or as end users using wireless communication services from terrestrial or aerial base stations [5]. For such mobile communication applications, uninterrupted UAV flights is critical to avoid any discontinuity in wireless communications, which may cause mission failure for some applications. Unfortunately, UAV flight-time is limited by the capacity of the on-board battery [1], raising a challenge for uninterrupted UAV missions, especially for smaller commercial drones.

On-going research to address UAV battery limitation largely concentrated on designing algorithms and motion control functions [6]–[11], that allow UAVs to operate in more energy-efficient way thereby extending the battery life. However, these efforts do not fundamentally solve the problem, because the UAVs still need to leave their missions and return to ground charging stations when the battery ultimately depletes. Utilizing the concept of far-field wireless power transfer (WPT), some researchers have recently contemplated the idea of charging the UAVs through the ground base stations, which eliminates the need for the UAVs to return to a charging station [12], [13]. However, the UAVs are required to remain in close proximity of the base stations during the WPT process. Efficient in-situ wireless charging of UAVs therefore remains a challenging open problem.

Inspired by the mid-air fueling of military jets using aerial tankers [14], in this paper we propose the concept of mid-air UAV wireless charging using specialized drones that are equipped with WPT equipment, henceforth referred to as tUAVs (transmitter UAVs). Under uncertain battery consumption in UAV-BSs due to dynamic mobile communication traffic, we then seek to optimize the trajectory of the flying tUAVs to maximize the long-term utility of the mobile communication service supported by the UAV-BSs. To the best of our knowledge, this is the first attempt to consider such wireless charging concept for UAVs using flying energy sources.

The main contributions of this paper can be summarized as follows: (i) we propose use of flying transmitter UAVs (tUAVs) to facilitate aerial wireless charging of UAV-BSs; (ii) we show that the trajectory optimization of the tUAVs can be modeled as an MDP given that the movement decision at any given step affects the long-term discounted utility of the underlying mobile communication system supported by the UAV-BSs; (iii) we solve the MDP problem using Q-learning and evaluate the performance of the proposed UAV charging architecture using simulations. Our results confirm that the optimized trajectory outperforms two benchmark strategies, namely random movement and static hovering of the tUAVs.

The rest of the paper is organized as follows. The related works are discussed in Section II. The problem statement and the proposed Q-learning-based trajectory optimization framework is presented in Section III. Performance evaluation of our proposed framework is discussed in Section IV. Finally, we conclude the paper and discuss future works in Section V.

[1]For example, the typical flight-time of *DJI Spreading Wings S900* drone is only 18 minutes when fully charged [4]

II. Related Work

Energy efficient UAV flight path planning, such as trajectory design methods were proposed in [6], [7]. While researchers optimize either the mechanical or the electronic energy consumption individually, work reported in [7] optimizes UAV trajectory to minimize the energy consumption for the mechanical as well as electronic functions of UAVs. Other researchers proposed a method to reduce the overall energy consumption of UAV communications by extending their network lifetime while guaranteeing their communication's success rate [8]. Optimal data collection techniques [9], [10] and a UAV-aided networking mechanism [11] are proposed which positively affect the UAV's energy consumption by optimizing their networking and communication methods.

Mechanisms for wireless recharging of UAVs have been proposed using terrestrial base stations using RF energy transmission in [12] and optical energy transmission in [13], however, the UAVs are required to stay close to some base stations in order to receive the power. This restricts the locations where the UAVs can be deployed, therefore, efficient in-situ wireless charging of UAVs remains a challenging open problem.

To address this gap, in our previous work [15], we introduced the use of aerial, stationary (i.e., hovering at fixed locations) RF chargers (tUAVs) placed at optimal locations with respect to the serving UAV-BSs (receiver UAVs, rUAVs) for maximizing the total delivered energy. The RF chargers are specialized UAVs carrying RF transmitters and hovering at the same height as the receiver UAV-BSs. We assumed that the tUAVs have sufficient power supply, e.g., by having a larger battery. The transmitter and receiver UAVs' altitude being the same, the placements of the transmitter UAVs were restricted to be outside the collision zone of two UAVs, i.e., the wingspan of separation. We studied their placement optimization to maximize the total harvested energy by the UAV-BSs from the received RF signals.

Since the received power is dependant on the distance the signals travel as per the well-known Frii's formula [16], having less or no restriction on the minimum distance between the tUAVs and the UAV-BSs compared to the wingspan distance used in [15], should enhance the received power levels. Moreover, the harvested energy should increase by having the tUAVs moving close to the UAV-BSs reducing the distance further, as opposed to a fixed optimal location. Therefore, building on the above proposal [15], in this paper we introduce the use of *mobile* RF chargers (tUAVs) that *fly and hover above* the rUAVs (as opposed to having a wingspan of separation distance at the same height of the rUAVs) and transmit wireless energy to extend their flying time. This also achieves direct LoS. The aim of this paper is to study the trajectories of such flying wireless chargers that would maximize the received power by the receiver UAVs. This architecture allows the chargers to carry any type of energy transmitters that can be used for wireless far field energy harvesting, e.g., RF omnidirectional antennae [17], massive MIMO with beamforming [18], Free Space Optics (FSO) [13], etc. In this paper, we assume highly directional RF antennae being carried by the tUAV as the source of the energy to increase the efficiency of WPT compared to omnidirectional antennae. We note that despite the specific type of wireless transmitter that is being used, the trajectory optimization of the flying chargers will hold applicable for any type of wireless transmitters due to the impact of distance and Line-of-Sight (LoS) requirement between the transmitter and receiver on the level of received power.

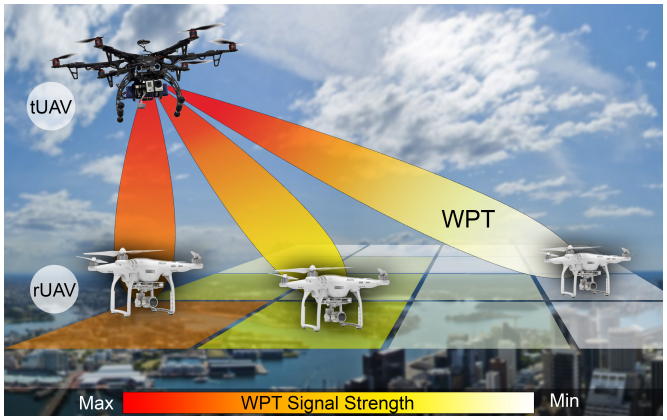

Fig. 1: In-situ recharging of UAVs, sample Wireless Power Transfer (WPT) beams in a flying trajectory.

III. System Description

In this section, we present our UAV recharging architecture, and the Q-Learning formulation to solve the trajectory optimization problem of the tUAVs.

A. UAV Recharging Architecture

Our proposed UAV recharging architecture consists of specialized, flying UAVs equipped with high gain RF antennae (tUAV) that transmit wireless power to in-situ recharge (or top-up) the batteries of UAV-BSs (rUAVs) that are deployed in an area to provide Hotspots for wireless communication to users/nodes. The architecture is illustrated in Figure 1. Our approach offers freedom to address the distance between the transmitters and receivers (tUAVs and rUAVs) that influences wireless energy transfer efficiency. The flying tUAVs are free to fly about and locate themselves such that they can minimize the distance and improve the line-of-sight RF links, as well as address the need of multiple rUAVs for transferring power. This increases the energy transfer utility and efficiency significantly. Unlike the mentioned related work on wireless recharging of UAVs, this architecture allows rUAVs to remain at their deployed locations or trajectories throughout their missions and provide their services while recharging. At the same time, the tUAVs may fly to new locations depending on the movement decision derived from the used algorithms. The decision is made based on rUAVs information which is sent to tUAV periodically. In this paper, we study the use of Q-Learning to make such decisions. Moreover, to formalize

the architecture as MDP, we assume that the geographical environment is discretized as a 2D grid where each cell can be covered by one rUAV. The tUAV and rUAVs are located at the center of the cells.

As we have used RF transmitters as the power source, we note that the received power of far-field RF transmission attenuates as per the reciprocal of the squared distance between the transmitter and the receiver. Therefore, assuming full energy conversion efficiency, the harvested RF power (P_r) at the rUAVs can be calculated using Frii's free space propagation model [16] as:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi d)^2} \quad (1)$$

where P_t is the transmit power, G_t and G_r are the antenna gains of the transmitter and the receiver, λ is the power transfer wavelength, and d is the distance between the transmitter and the receiver.

In particular, we consider that the tUAV is equipped with a few high gain antennae to transfer energy to rUAVs. Since the locations of the UAVs are not being changed rapidly, we assume that each of the tUAV's antennae can be aligned mechanically toward one of the rUAVs. The antenna alignment and high gain can be achieved by MIMO beamforming as an alternative which is not in this paper scope, and left for our future work. Also, a small amount of data is communicated between tUAV and rUAVs periodically using the same wireless channel. This data includes location, battery status and consumed-received energy of each rUAV. Since the RF energy harvesting is limited and may not provide infinite flying time for rUAVs, we assume that the low battery rUAV leaves the network to be charged with other wired or near field WPT methods. However, our solution aims to keep them in operation for as long as possible.

In this paper, our proposed charging technology is limited to RF power transfer, but the power transfer can be via laser beam and also rUAVs can exploit other in-situ battery charging techniques such as solar energy. We will consider hybrid charging in our future works. We expect our Q-learning approach to be compatible with all of these in-situ charging methods.

B. Q-Learning Formulation

Q-learning [19] is a model-free reinforcement learning algorithm in which an agent transitions from one state to another, by taking random actions. A set of states S and set of actions A define the learning space. By performing an action $a \in A$ and moving to another state, a *revenue* function calculates a numeric value for taking such state-action pair and records it in a Q-table which is initialized with zero values. By repeatedly taking random actions, at one point the agent reaches a particular goal state. The Q-table values get updated at each step and after many iterations eventually converge. Q-Learning's goal is to maximize the total revenues for all state-action pairs from beginning up to reaching the goal state, so called the *optimal policy*. The optimal policy indicates which action is the best to take in different states, which results in a maximized overall gain.

Q-Learning has been widely used in UAV related research recently. This includes a range of application from military threat avoidance [20] and obstacle avoidance [21] to trajectory optimization for improving services in wireless communications [22], [23].

In this paper, we employ Q-learning to find the best location and movement for the tUAV at a given observation of entire network of rUAVs. Q-Learning components in our solution are defined as following:

- **Agent** (tUAV) observes the current state and takes actions to move to other states.
- **Action** (a) is defined as flying to a neighboring cell in our assumed grid space, or hovering over the current cell. Hence, as shown in Figure 2, we consider 9 possible actions. Some actions are not available on the edges of our considered area.
- **State** (S) is defined based on the observed information of rUAVs and the current location of tUAV. Thus, we define the state as $S = \{L_c, L_h, B_h\}$ where L_c is the location of tUAV, $L_h = [L_{h_1}, L_{h_2}, ..., L_{h_n}]$ is a vector that denotes the location of rUAV1 to rUAVn and $B_h = [B_{h_1}, B_{h_2}, ..., B_{h_n}]$ is a vector that denotes the battery level of rUAV1 to rUAVn.
- **Revenue** (R) is a function of last state and action which returns a reward for the energy that all rUAVs receive from tUAV and/or applies a penalty if an rUAV has to move to terrestrial charging station because of low battery. R is formulated as:

$$R(S, a) = \mu E_{r,tot} + \nu N_o \quad (2)$$

where $E_{r,tot}$ is the total harvested energy by rUAVs and N_o is the number of out of charge rUAVs. μ and ν are adjusting factors.

We consider a time step, T in our model which is roughly long enough for the tUAV to fly from one cell to another. This time step is used to take actions over time, and update the Q-Table values after completing every transition and observing the $current$ state-action pair (s, a) as:

$$Q^{new}(s, a) = (1 - \alpha)Q(s, a) + \alpha(R(s, a) + \gamma Q(s', a^*) \quad (3)$$

where α and γ are learning rate and discount factor respectively, s' is the next state after taking action a at state s and a^* is the action that results in the maximum Q-value of all state-action pairs on state s':

$$a^* = argmax_a Q(s', a) \quad (4)$$

In the above model, the agent (tUAV) needs to observe the rUAVs' geographical locations and their remaining battery levels. We assume our rUAVs remain in the same geo-cell in our considered area, therefore, only their battery status needs to be sent to the tUAV at each time step. Hence, our tUAV

and rUAVs must have a light periodic signalling to exchange information.

Considering the discussed Q-Learning components, we follow Algorithm 1, to obtain an optimal flying trajectory and recharging mechanism that maximizes the overall flying duration of all rUAVs. In this algorithm, the tUAV receives updated information of the rUAVs at each time step. The observed data including the tUAV current location indicate the current state. The agent makes a decision on movement either randomly or based on the Q-Learning policy. We employed ϵ-greedy scheme for exploration in Algorithm 1 where the probability of selecting a random action decreases while the policy is being optimized through the iterations.

Algorithm 1 : Q-Learning

```
Initialize Q-Table to zero
Initialize tUAV location
Set Exploration more than 1
Observe rUAVs locations
REPEAT
  Observe rUAVs Battery
  Current state = (tUAV location,Observation)
  Generate a random number r ∈[0,1]
  IF r is less than Exploration
    Update tUAV location by a random action
  ELSE
    Update tUAV location by the best action for current state
  Calculate Revenue(consumed-received energy, out of charge rUAVs)
  Update Q-Table by Revenue
  Decrease Exploration
CONTINUE
```

IV. Performance Evaluation

A. Simulation Setup

In our scenarios, we consider three rUAVs and one tUAV, located in an environment modeled as a 3 by 3 grid (Figure 2, 3). To simplify our simulation design, we assume all tUAV and rUAVs can be located only at the center of these geo-cells. The tUAV sends the charging signal to all rUAVs at the same time and the closest rUAV receives the most amount of RF energy. Figure 2 demonstrates possible flying directions (i.e., actions) of our tUAV when it's located in different geo-cells. Figure 3 illustrates three different scenarios that were considered in our simulations. For each scenario, we run our simulated Q-Learning model for 50000 episodes (i.e., iterations). In each episode, the tUAV starting cell is random. Also, the battery of each rUAV is initialized randomly between 60 and 100 Watt-hour. The battery level is discretized to 5 levels as an observable parameter for Q-Learning. In the first episode, the exploration rate of Q-Learning is 1 and the trajectory of the tUAV is completely random. The tUAV location is updated at the start of each time step and the amount of consumed and received energy by each rUAV is also calculated at the end of each period. Consequently, the Q-Table is updated using equation (3) when the new battery levels are observed. Since the harvested RF energy is less than rUAV energy consumption, rUAVs will be out of charge eventually. The episode continues until all rUAV batteries are discharged and they have to fly back to the terrestrial charging station. The exploration rate decreases linearly to zero after around 40000 episodes and then, the tUAV trajectory is fully based on Q-Table policy where at each observed state, the action with maximum Q-value is taken as the optimum decision using equation (4). The simulation parameters are listed in Table I.

In order to evaluate our algorithm's performance, we consider the level of delivered wireless energy to the rUAVs and also simulate the following two baseline tUAV trajectory models as benchmark algorithms:

- Random trajectory: all movements of the tUAV are random. Staying in the same geo-cell is also allowed in this model.
- Static hovering: The tUAV hovers in the central geo-cell for the entire episode.

TABLE I: Simulation Parameters

Q-Learning Component	Value
Transmit power	35 Watt
Antenna gain	25 dBi
Cell side	10 m
Charging Wave Frequency	25 GHz
Learning rate	0.4
Discount factor	0.95
rUAV power consumption	50 Watt
rUAV battery capacity	100 Watt-hour
Time step	20 Sec
Revenue adjusting factor (μ)	100
Revenue adjusting factor (ν)	-50

B. Results

To demonstrate the exploration progress of Q-Learning Agent, the average flying time of rUAVs for three scenarios are shown in Figure 4. This illustrates how the learning is progressing when at the start of the learning process, our Q-learning model takes totally random actions and gradually decreases this randomness by taking the Q-Table values into account. After around 40000 episodes, when the exploration rate drops to zero, the tUAV has a policy to find the best trajectory based on rUAV locations and their battery status. We find that flying time is increased from 71.5 to 77 minutes during the training session.

In Figure 5, we compare how the Q-Learning based trajectory of the tUAV performs against that of two benchmark trajectories in terms of resulting average flying time of the rUAVs. As it is shown, the Q-Learning model significantly outperforms both methods since the tUAV makes decisions based on the knowledge obtained from previous iterations and aims to maximize the transferred energy and extend flying time of the rUAVs. The standard error bars also are shown on the chart which are mostly influenced by battery level randomness of each episode and are observed to be close for all schemes.

The energy transfer efficiency is limited mainly by distance and to increase the received energy, a more powerful electromagnetic wave can be emitted towards receivers. We repeat the Scenario-1's simulation for various transmission power.

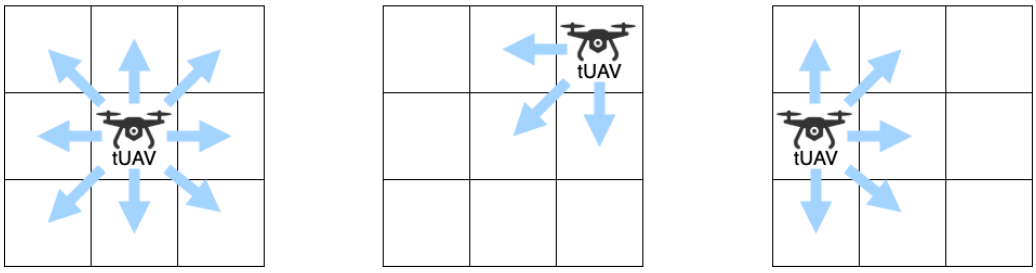

Fig. 2: Possible movement directions of the transmitter UAVs (tUAV) with respect to its three example current locations. The tUAV periodically changes its position to improve the energy transfer efficiency. In some periods, the tUAV may find that its current location is the best, thus the position may not be changed.

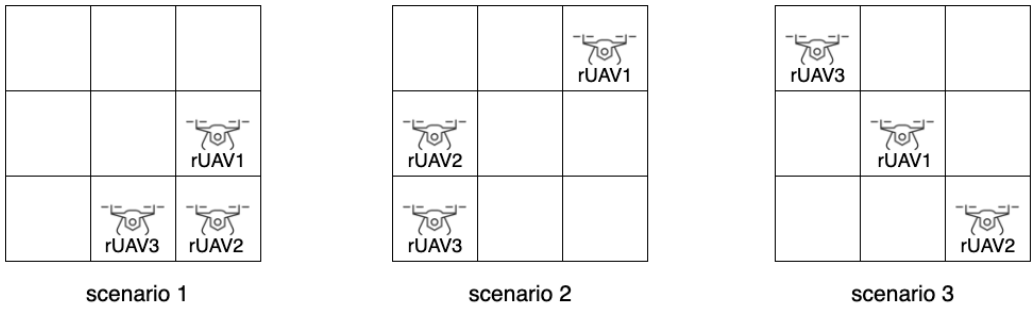

Fig. 3: Considered simulation scenarios showing the positions of the rUAVs. The rUAVs are stationary in each scenario where the initial position of tUAV is randomized for each episode.

Figure 6 presents the average results. As it is illustrated, our Q-Learning based model is more effective than baseline positioning models when a higher amount of energy is emitted. For example, when the transmit power is 55 W, Q-Learning has an average flying time gain by 14% and 19% in comparison with the random trajectory and fixed hovering location of the tUAV at center of the map respectively. Currently, we have used a single tUAV in our solution, however, supporting multiple tUAVs would also enhance the level of received energy. This will require a multi-agent Q-Learning formulation. We will explore this multi-agent approach that supports multiple tUAVs in our future work.

V. Discussion and Future Work

We introduced the concept of using *mobile, aerial* chargers for in-situ topping up of Hotspot UAV-BS battery using wireless power transfer. In order to enhance the level of received wireless power, we formulated the trajectory optimization problem of the aerial charger using MDP, and solved it using Q-Learning to maximize the flying time of energy thirsty UAV-BSs. Using simulation studies, we demonstrated that the Q-Learning based optimized trajectory of the aerial charger outperforms the benchmarking trajectories. Although our solution targeted Hotspot UAV-BSs that hover above fixed locations, it can be generalized for all applications where the power receiver UAVs (rUAVs) are hovering. Future work could consider other applications with mobile rUAVs.

The Q-Learning is a dynamic programming method which can update the decision policy in an adaptive manner. In

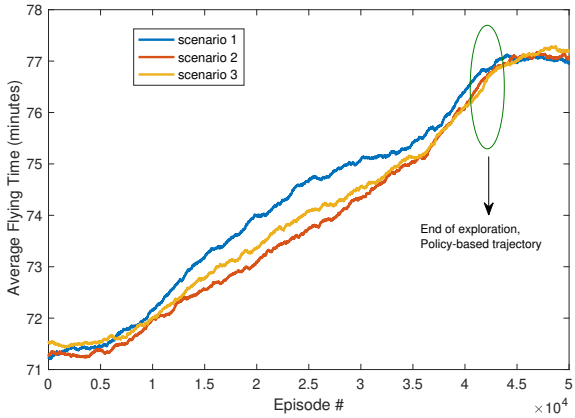

Fig. 4: The average flying time of rUAV. The exploration rate is decreasing to the point that labeled on graphs.

situations where the policy requires many experiences initially to achieve an acceptable performance, this training or exploration period can be estimated offline and loaded in the agent's software to avoid poor performance during the training phase. Thereafter, the agent which is the flying energy source (tUAV) in our studied problem, updates the policy using real environmental observations and experiences.

To avoid complexity, we chose a small scale scenario to demonstrate the usefulness of Q-Learning in the placement optimizations of the flying energy sources. However, a real

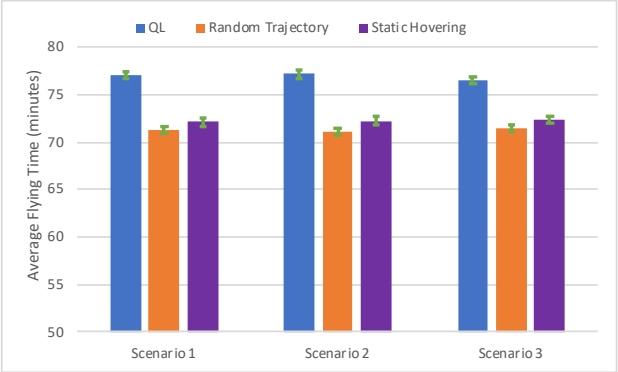

Fig. 5: Comparison of flying time extension by positioning of the tUAV using Q-Learning and naive baselines.

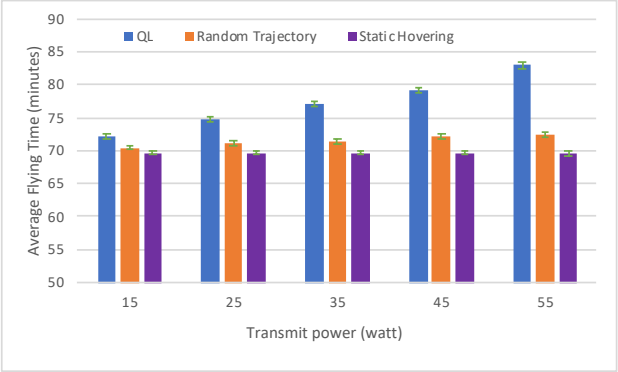

Fig. 6: Comparison of flying time extension by positioning of the tUAV using Q-Learning and naive baselines for different transmit power.

scenario may be larger than what we have studied, where tens or hundreds of flying UAVs may [24] require in-situ charging in a large area and by multiple tUAVs. In such scenarios, basic Q-Learning may be too naive to meet the requirements and be practically implementable. For this purpose, more complex variations of Q-Learning need to be considered. In our future work, we will focus on methods such as Deep Reinforcement Learning [25], where the neural network is employed to support a Deep Q Network (DQN). This can support multi-agents and continuous values for observations and actions, while supporting a large scale environment.

ACKNOWLEDGEMENT

This work is supported by the Central Queensland University (CQU) Research Grant RSH5137.

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] M. Mazur, A. Wiśniewski, R. Abadie, V. Huff, J. Smith, and S. Stroh, "Clarity from above," 2019, accessed = 2019-12-2. [Online]. Available: https://www.pwc.pl/clarityfromabove

[3] A. Fotouhi, M. Ding, and M. Hassan, "Flying drone base stations for macro hotspots," *IEEE Access*, vol. PP, pp. 1–1, 03 2018.

[4] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L. G. Giordano, A. Garcia-Rodriguez, and J. Yuan, "Survey on uav cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, 2019.

[5] R. Zhang and Y. Zheng, "Accessing from the sky: Uav communications for 5g and beyonds (tutorial)," in *IEEE International Conference on Communications (ICC) (Tutorial)*. IEEE, 2019.

[6] D.-H. Tran, T. X. Vu, S. Chatzinotas, S. ShahbazPanahi, and B. Ottersten, "Trajectory design for energy minimization in uav-enabled wireless communications with latency constraints," *arXiv preprint arXiv:1910.08612*, 2019.

[7] S. Salehi, A. Bokani, J. Hassan, and S. S. Kanhere, "AETD: An Application Aware, Energy Efficient Trajectory Design for Flying Base Stations," in *2019 IEEE 14th Malaysia International Conference on Communication (MICC)*, December 2019.

[8] 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*, vol. 15, no. 6, pp. 1377–1386, 2015.

[9] A. E. Abdulla, Z. M. Fadlullah, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "An optimal data collection technique for improved utility in uas-aided networks," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 736–744.

[10] C. Zhan, Y. Zeng, and R. Zhang, "Energy-efficient data collection in uav enabled wireless sensor network," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 328–331, 2017.

[11] A. E. Abdulla, Z. M. Fadlullah, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "Toward fair maximization of energy efficiency in multiple uas-aided networks: A game-theoretic methodology," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 305–316, 2014.

[12] M. Hua, C. Li, Y. Huang, and L. Yang, "Throughput Maximization for UAV-enabled Wireless Power Transfer in Relaying System," in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2017, pp. 1–5.

[13] N. Ansari, D. Wu, and X. Sun, "FSO as backhaul and energizer for drone-assisted mobile access networks," *ICT Express*, 2020.

[14] R. K. Nangia, "'Greener' civil aviation using air-to-air refuelling – relating aircraft design efficiency and tanker offload efficiency," *The Aeronautical Journal (1968)*, vol. 111, no. 1123, p. 589–592, 2007.

[15] J. Hassan, A. Bokani, and S. S. Kanhere, "Recharging of flying base stations using airborne rf energy sources," in *2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, April 2019, pp. 1–6.

[16] "The Friis Equation," http://www.antenna-theory.com/basics/friis.php, 2009-2015 (accessed January 15, 2019).

[17] M. Arrawatia, M. S. Baghini, and G. Kumar, "Broadband bent triangular omnidirectional antenna for rf energy harvesting," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 36–39, 2015.

[18] Y. Wang, A. Liu, K. Xu, and X. Xia, "Energy and information beamforming in airborne massive mimo system for wireless powered communications," *Sensors*, vol. 18, no. 10, p. 3540, 2018.

[19] C. J. C. H. Watkins and P. Dayan, "Q-learning," in *Machine Learning*, 1992, pp. 279–292.

[20] C. Yan, X. Xiang, and C. Wang, "Towards real-time path planning through deep reinforcement learning for a uav in dynamic environments," *Journal of Intelligent & Robotic Systems*, Sep 2019.

[21] Z. Yijing, Z. Zheng, Z. Xiaoyi, and L. Yang, "Q learning algorithm based uav path learning and obstacle avoidence approach," in *2017 36th Chinese Control Conference (CCC)*. IEEE, 2017, pp. 3397–3402.

[22] U. Challita, W. Saad, and C. Bettstetter, "Deep reinforcement learning for interference-aware path planning of cellular-connected uavs," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–7.

[23] H. Bayerlein, R. Gangula, and D. Gesbert, "Learning to rest: A q-learning approach to flying base station trajectory design with landing spots," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 724–728.

[24] A. Agogino, C. HolmesParker, and K. Tumer, "Evolving large scale uav communication system," in *Proceedings of the 14th annual conference on Genetic and evolutionary computation*. ACM, 2012, pp. 1023–1030.

[25] L.-J. Lin, "Reinforcement learning for robots using neural networks," Carnegie-Mellon Univ Pittsburgh PA School of Computer Science, Tech. Rep., 1993.