

Article

# In Situ MIMO-WPT Recharging of UAVs Using Intelligent Flying Energy Sources

Sayed Amir Hoseini <sup>1</sup>, Jahan Hassan <sup>2,\*</sup>, Ayub Bokani <sup>2</sup> and Salil S. Kanhere <sup>3</sup>

<sup>1</sup> School of Engineering and Information Technology, The University of New South Wales, Canberra, ACT 2600, Australia; s.a.hoseini@unsw.edu.au

<sup>2</sup> School of Engineering and Technology, The Central Queensland University, Sydney, NSW 2000, Australia; a.bokani@cqu.edu.au

<sup>3</sup> School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 2052, Australia; salil.kanhere@unsw.edu.au

\* Correspondence: j.hassan@cqu.edu.au

**Abstract:** Unmanned Aerial Vehicles (UAVs), used in civilian applications such as emergency medical deliveries, precision agriculture, wireless communication provisioning, etc., face the challenge of limited flight time due to their reliance on the on-board battery. Therefore, developing efficient mechanisms for in situ power transfer to recharge UAV batteries holds potential to extend their mission time. In this paper, we study the use of the far-field wireless power transfer (WPT) technique from specialized, transmitter UAVs (tUAVs) carrying Multiple Input Multiple Output (MIMO) antennas for transferring wireless power to receiver UAVs (rUAVs) in a mission. The tUAVs can fly and adjust their distance to the rUAVs to maximize energy transfer gain. The use of MIMO antennas further boosts the energy reception by narrowing the energy beam toward the rUAVs. The complexity of their dynamic operating environment increases with the growing number of tUAVs and rUAVs with varying levels of energy consumption and residual power. We propose an intelligent trajectory selection algorithm for the tUAVs based on a deep reinforcement learning model called Proximal Policy Optimization (PPO) to optimize the energy transfer gain. The simulation results demonstrate that the PPO-based system achieves about a tenfold increase in flight time for a set of realistic transmit power, distance, sub-band number and antenna numbers. Further, PPO outperforms the benchmark movement strategies of “Traveling Salesman Problem” and “Low Battery First” when used by the tUAVs.

**Keywords:** UAVs; wireless power transfer; RF energy harvesting; MIMO; deep reinforcement learning



**Citation:** Hoseini, S.A.; Hassan, J.; Bokani, A.; Kanhere, S.S. In Situ MIMO-WPT Recharging of UAVs Using Intelligent Flying Energy Sources. *Drones* **2021**, *5*, 89. <https://doi.org/10.3390/drones5030089>

Academic Editors: Higinio González Jorge and Luis Miguel González de Santos

Received: 12 July 2021

Accepted: 1 September 2021

Published: 5 September 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The recent years have seen increasing advancements and decreasing costs of low-altitude UAVs, commonly known as drones. Drones carrying a range of technologies for sensing and communication are becoming popular with service providers as innovative service delivery platforms, such as for emergency medical deliveries, precision agriculture, aerial imagery, etc. This popularity contributes directly to the growth of the global market for drone-delivered commercial services to an estimated value of *USD127bn* [1]. Drones are also employed in 5G networks either as aerial base stations providing a wireless Hotspot or mobile relaying services to the ground nodes [2,3], or as aerial nodes of cellular UAV networks [4,5].

With such a staggering market value, reliability through service continuity becomes a critical success factor [6]. However, drones have short flying times due to their dependency on on-board, limited capacity batteries for power supply. For example, the typical flight-time of the DJI Spreading Wings S900 drone is about 18 min when the battery is fully charged [3]. This implies that drones need to make frequent trips to the ground charging stations so their batteries can be replaced or recharged, which creates significant service

disruptions. To reduce the disruption, in situ recharging of the drone battery using ambient energy harvesting techniques is considered as a core technology for operational UAVs in 5G networks [6].

The energy limitation issue is largely being addressed through the design and optimization of algorithms and motion control functions [7–12] to achieve energy efficiency. While such efforts are helpful, they do not fundamentally solve the problem since the drones would still need to fly away from their missions and return to ground charging stations when the battery eventually drains out. Solar powered drones can harvest energy from the sun. To harvest enough energy, the drones need fixed wings with a long wingspan (e.g., 4 m [13]) to accommodate the solar panels. As such, smaller, consumer drones as well as rotary-wing drones cannot benefit from this solution. Moreover, the solar energy harvesting is dependent on flight conditions, e.g., cloudy days and night time are not favorable for this type of energy harvesting. The far-field WPT using the Radio Frequency (RF) Electromagnetic Radiation (EMR) technique (radiative WPT) is a promising approach for powering UAVs [14,15], which allows the transmitter and receiver to be located over a distance. This specifically suits the deployment and mobility requirements of recharging UAVs during missions.

The viability of the far-field WPT approach was demonstrated in [16–19]. William C. Brown for instance, showed how a wireless-powered helicopter can be powered over a distance of 18 m above the transmitting antenna in 1964 [16]. In this experiment, 270 W power could be harvested at 2.45 GHz. Recently, the simultaneous wireless information and power transfer system was explored to send wireless information and power to drones from terrestrial base stations [20,21]. However, the drones need to remain in the close proximity to the base stations during the WPT process to achieve Line-of-Sight (LoS) links. This limits the deployment location and mobility of the UAVs. Other issues with far-field WPT are the drop in transmission efficiency due to high path loss when the distance between the transmitter and receiver is increased [15] and the random, uncontrollable energy arrival at the receiver when non-dedicated energy sources are used [6].

In this paper, therefore, we propose the deployment of multiple, flying energy transmitters for recharging UAVs (rUAVs) using WPT. This is inspired by the practice of mid-air fueling of military jets using aerial tankers, a concept that was also proposed for civil aviation purposes [22]. The transmitters are specialized UAVs (tUAVs) equipped with Multiple-Input-Multiple-Output (MIMO) antennas. The MIMO antenna system can direct energy beams towards the receivers. The rUAVs can continue their mission without having to adjust their locations to receive power, while the tUAVs dynamically adjust their locations to reduce the transmission distance to enhance the power delivery. To boost the WPT efficiency while meeting the regulatory constraints on maximum transmission power [23], in each tUAV, we propose deploying multiple antennas that operate in multiple band spectrum. This is because the use of multi-band transmission helps to distribute power over a wider spectrum so that the maximum transmitted power does not exceed the regulated limit. The rUAVs convert the received RF power to DC power using an array of rectennas that are special types of antennas used to convert electromagnetic energy to DC current. A conceptual view of our proposal is shown in Figure 1.

The energy consumption of each rUAV may be different due to the environmental conditions (e.g., windy or still conditions), dynamic wireless communication requirements, and mobility. Therefore, the tUAVs must intelligently pick the rUAVs to serve according to their residual energy levels in a coordinated manner. This calls for a multi-agent optimization model, for which we employ Proximal Policy Optimization (PPO), a recent class of Deep Reinforcement Learning (DRL) algorithms. PPO adjusts each tUAV's movements to intelligently pick the next rUAVs to be recharged considering traveling time, other rUAV locations, rUAV battery level and the other tUAVs' locations. This minimizes the service interruptions by extending flying times of the rUAVs. In other words, using PPO, each tUAV can find the best location to move to at a given observation of entire network of rUAVs.

To the best of our knowledge, this is the first attempt to consider such a wireless charging architecture for UAVs using multiple dedicated and coordinating aerial energy sources.

The main contributions of this paper can be summarized as: (i) we propose a system of multiple tUAVs to facilitate aerial wireless charging of rUAVs using multi-band MIMO beamforming, (ii) we propose a PPO-based movement decision algorithm for the tUAVs in selecting the next rUAVs to recharge as per their battery levels, and (iii) we compare the PPO-based system performance using simulations, with two benchmark movement decisions strategies of the tUAVs: traveling Salesman Problem (TSP) and Low Battery First (LBF). Our results demonstrate that with PPO, the system achieved a tenfold flight time extension compared to no WPT. Further, this strategy outperforms the benchmark movement strategies of TSP and LBF when used with WPT.

The rest of the paper is organized as follows. The related works are discussed in Section 2. System description and the DRL model is presented in Section 3 followed by the performance evaluation of proposed model in Section 4. We finally conclude the paper and discuss future works in Section 5.



**Figure 1.** In situ recharging of UAVs using aerial wireless energy sources: Sample Wireless Power Transfer (WPT) beams in a flying trajectory.

## 2. Related Works

Researchers commonly address the drone energy-limitation issue through the design of their energy-efficient functioning mechanisms. These mechanisms include flight path (trajectory) planning and communication methods. UAVs consume energy due to their mechanical (flying, hovering) and electronic (wireless communication) functions, presenting scopes for improving energy efficiency of both. However, since the mechanical energy consumption is significantly more than that from the electronic functions, researchers mostly focus on optimizing the trajectory to shorten the flight paths for reducing mechanical energy consumption, e.g., in [7,8,24]. As our current work is on the topic of energy replenishment solutions for the deployed UAVs, we omit details of the work on energy-efficient UAV operations.

One approach to address the energy replenishment objective of the deployed UAVs is using tethered UAVs [25], wherein the UAVs are connected via a cable to the ground station to receive continuous supply of power. As the ground station has an unlimited power supply, the tethered UAVs can operate perpetually. However, this approach restricts the UAVs' mobility and deployment locations to be only in the areas with existing ground stations. Another solution is to use UAV battery swapping [26], which requires the deployed UAVs to fly back to the ground station for "hot-swapping" the battery, whereby the external power sources keep the UAVs powered on while the battery replacement takes place. This approach is quicker than recharging the on-board battery at the ground station. However, this approach leads to service interruptions due to the UAVs leaving the serving area for the battery swapping. To allow greater freedom of deployment locations and

mobility of the UAVs, we consider a far-field wireless powering approach in our current work for which we discuss the related work below.

As previously mentioned, there have been many trials in the past demonstrating the viability of far-field WPT techniques. Further development and practical use of such techniques are somewhat stalled; however, we see a renewed interest, as evident in recent industry needs, activities and trials. A New-Zealand based startup company, Emrod, was recently reported to have developed a long-range, high-power, WPT technology to deliver wireless electricity to end users without needing copper power lines [27]. This follows the country's second largest power company Powerco, planning to trial the technology in 2021 [28]. In 2020, a US-based company, Powerlight Technologies (formerly known as Laser-Motive), demonstrated a wireless power receiver for drones. In an earlier demonstration, the company used a laser beam to fly a drone for more than 12 h [29].

On the other hand, the increasing demand for UAVs' autonomous wireless recharging is clearly evident by their numerous commercial operations such as power line monitoring (e.g., [30]), food delivery (e.g., [31]) and law enforcement (e.g., [32]), to name a few. The short flight times of these drones are causing serious deployment hindrances in these industries. According to a 2020 Bloomberg report, the predicted market value of the global autonomous wireless charging and infrastructure market for drones will reach USD 249.3 Million by 2024 [33]. This justifies the unified efforts from research and industry that are required to develop practical solutions for wireless charging of drones and, more importantly, in situ solutions.

The utility of far-field WPT using EM radiation is well established: it provides placement flexibility and mobility of transmitters and receivers, can work even in non-LoS conditions, and can power over a distance. Due to the energy conversion efficiency limitation, generally this technique is suitable for low-power devices. Despite this limitation, various works on WPT suggest that far-field WPT can also be used for recharging UAV batteries (e.g., [15]). To this end, wireless recharging of UAVs were proposed using RF WPT in [20] and optical energy transfer in [21], both from ground base stations in a simultaneous wireless information and power transfer system. Using a power-splitting and time-switching architecture, authors proposed a relaying system in [20] in which the UAV harvests energy and information from the base station, and relay the information to a ground node. With an objective of prolonging the lifetime and throughput maximization of the network, authors optimized the system parameters along with the UAV deployment location; however, no explicit results on received power were mentioned. The authors in [21] studied a similar system but with an optical transmitter at the ground base station casting optical beam to the UAV carrying both data and energy, providing simultaneous communication and charging. The numerical results showed that the system achieved a high network throughput and a 25% extra hovering time in the drones. However, both proposals require the UAVs to be in the proximity of the terrestrial base station to achieve LoS and receive power. This limits the locations where the UAVs can be deployed due to the fixed terrestrial base stations. Therefore, flexible in situ wireless charging of UAVs remains a challenging open problem.

In our previous works [34,35], we studied different modes of dedicated, aerial WPT chargers to observe their performance subsequently, with tUAVs carrying omnidirectional antennas. In [34] we utilized aerial, stationary (i.e., hovering at fixed locations) tUAVs to study their optimal placement locations with respect to the rUAVs to maximize total received power at the rUAVs. In [35], we utilized one flying tUAV to power all rUAVs. In that work, for the single tUAV to recharge all rUAVs, a single-agent optimization of the tUAV's trajectory was presented via Q-Learning to enhance power delivery. However, Q-Learning comes with a scalability issue and both observation and action space must be limited. As such, for multi-agent systems (current work) Q-learning poses limitations. Further, the use of omnidirectional antennas waste energy since energy is radiated all around the antenna, not only to the energy receiver. The fundamental differences of our current work with our prior works is that our current work uses multi-band MIMO

antennas at the flying energy sources which distributes power over a wider spectrum and exploits targeted energy beams through beamforming to recharge chosen UAVs. This boosts energy delivery at the rUAVs. Further, we employ a multi-agent optimization model for multiple tUAVs in the network to optimize the tUAVs' movement decisions using the PPO algorithm. The advantages of PPO over Q-Learning is discussed in the next section.

### 3. System Description

In this section, we present our UAV recharging architecture involving multiple tUAVs and rUAVs. We also present the deep reinforcement learning algorithm using the PPO technique to control the movements of the tUAVs in targeting the next rUAVs to recharge. The optimization aims to enhance the MIMO-WPT efficiency to achieve longer flying times of the rUAVs in the presence of multiple coordinated tUAVs serving multiple rUAVs with dynamic battery levels.

#### 3.1. UAVs Recharging Architecture

Our proposed UAV recharging architecture consists of specialized, flying UAVs equipped with multiple high gain RF antennas (tUAV) that transmit wireless power to recharge the rUAVs' batteries. We assume that the rUAVs are deployed in an area to provide Hotspot wireless communication services to the ground users (Figure 1). The tUAVs are assumed to have a significantly greater power supply than the commodity rUAVs, e.g., by carrying a larger battery or having hybrid power sources. As such, the tUAVs are expected to be costlier and bulkier than the rUAVs. Further, a tUAV which recharges several rUAVs can be replaced with another tUAV when its energy is depleted. However, the tUAV replacement does not interrupt the services of the rUAVs. It is to be noted that our aim is to extend each rUAV's operating time as much as possible, thus reducing the number of times the rUAVs would return to the ground station once their batteries eventually deplete. The tUAVs fly about and position themselves in a way to minimize the distance between them and the rUAVs, and to improve the line-of-sight RF links for the target rUAVs. This enhances the power transfer effectiveness.

To increase energy transfer efficiency, we propose a MIMO system to perform an energy beamforming and focus energy toward the receiver [36–38]. Hence, we consider a point-to-point MIMO system with  $m_t$  antennas installed on the tUAVs and  $m_r$  antennas on the rUAVs. Without the loss of generality, we assume a uniform square array of antenna on each side. We use the system model obtained from [14,37] where a total of  $N$  orthogonal sub-bands are used to transmit energy. On each sub-band, sine-wave signal  $\mathbf{S}_n$  is emitted at carrier frequency  $f_n$  by  $m_t$  tUAV antennas as

$$\mathbf{s}_n(t) = [s_{1n}(t), \dots, s_{m_t n}(t)]^T, \quad (1)$$

where  $n = 1, 2, \dots, N$  and  $s_{mn}(t)$  is the beamforming component of  $s_n(t)$  by antenna  $m$  at frequency  $n$ . Thus, the total received power at all  $m_r$  receiver antennas is defined as

$$P_r = \sum_{i=1}^{m_r} \sum_{n=1}^N \mathbb{E} [|\mathbf{h}_{in}^H \mathbf{s}_n(t)|^2] = \sum_{n=1}^N \text{tr}(\mathbf{H}_n^H \mathbf{H}_n \mathbf{S}_n),$$

where  $\mathbf{h}_{in}^H [h_{i1n}^*, \dots, h_{im_t n}^*]$  is the channel vector from transmitter antennas to receiver antenna  $i$ ,  $\mathbf{H}_n$  is the channel matrix between  $m_t$  transmitter antennas and  $m_r$  receiver antennas and  $\mathbf{S}_n$  is the transmit covariance matrix all at sub-band  $n$ . Similarly, the total transmit power at frequency  $f_n$  is

$$P_t = \sum_{n=1}^N \text{tr}(\mathbf{S}_n). \quad (2)$$

The maximum transmit power at each sub-band is constrained by regulation and hardware limits. Thus, let us assume

$$\text{tr}(\mathbf{S}_n) \leq P_s, \forall n. \quad (3)$$

Based on [14] and assuming the maximum sum-power  $P_s$  is transmitted at each sub-band, the received power at each sub-band  $n$  is obtained as

$$P_{r,n} = P_s \lambda_{\max,n}^2, \quad n = 1, \dots, N, \quad (4)$$

where  $\lambda_{\max,n} = \lambda_{\max}(\mathbf{H}_n^H \mathbf{H}_n)$  denotes the maximum singular value of  $\mathbf{H}_n^H \mathbf{H}_n$  for sub-band  $n$ . As a result, the total harvested energy is

$$P_r = P_s \sum_{n=1}^N \lambda_{\max,n}^2. \quad (5)$$

Since there is almost a pure LoS MIMO channel between a pair of tUAV and rUAV during the recharging process as the tUAV adjusts its position to achieve this,  $\mathbf{H}_n$  is a rank one matrix and an optimal beamforming can be achieved by an SVD-based beam-former [39,40] when only one strong beam is formed by transmitter antennas as *optimal energy beamforming* with a gain of  $n_t$  [37,41],

$$\lambda_{\max,n} = a_n \sqrt{m_t m_r}, \quad (6)$$

where  $a_n$  is the signal attenuation along the LoS path at frequency  $n$  which is assumed to be the same for all antenna pairs. This assumption is valid when the distance between transmitter and receiver is much larger than the antenna array size [39]. For this purpose, the Channel State Information (CSI) should be available at transmitter side. In contrast with MIMO channel's information, the energy transfer channel in our system is significant, stable and relatively time invariant. Hence, measuring the CSI feedback is not a challenging task. Attenuation is also achieved by

$$a_n^2 = \frac{G_t G_r c^2}{(4\pi d f_n)^2} \quad (7)$$

where  $G_t$  and  $G_r$  represent each antenna gain at the transmitter and receiver, respectively,  $c$  is the speed of light and  $d$  is the distance between the transmitter and receiver. Thanks to mechanical alignment, high gain antennas can be employed to boost MIMO gain [41,42]. We applied a limit of 90% efficiency [43] to the RF gain in (5) to model nonideal implementation of MIMO system, i.e., mutual coupling. There is also RF to DC conversion efficiency at the receiver which represents how much of received wireless energy can be converted to usable energy by the rUAVs. In this work, we assume a constant RF to DC efficiency of 80% [44,45] that is denoted by  $\gamma$ . Additionally, we assume that the receiver antennas are installed on the top of the rUAVs to minimize blockage by rUAV's frame or blades. Equation (7) shows that the energy transmission is significant for short distances; therefore, the tUAV should hover above rUAV to maximize energy transfer and this will minimize the blockage. Furthermore, small movements of both tUAV and rUAV do not reduce efficiency of beam alignment since CSI can be measured several times in a second to update beam direction. As discussed earlier, we limited the RF efficiency in our proposed conjecture to model non-ideal implementation. However, the true and highly accurate model can be achieved by a prototype system.

### 3.2. Proposed Trajectory Selection Algorithm

Proximal policy optimization (PPO) is a model-free, online, on-policy reinforcement learning method from policy gradient family [46,47]. This method supports both discrete and continuous spaces for observations and actions. A PPO agent transitions from one state to another, by taking random actions. A set of states  $S$  which are defined based on the

observations from the environment and a set of actions  $A$  define the learning space. By performing an action  $a \in A$  and observing the resulting state, a revenue function calculates a numeric reward. The learner's goal is to maximize the discounted long-term reward 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. PPO selects actions based on the probability distribution and we define the optimal policy so that the action with maximum likelihood is chosen after training as deterministic exploitation. PPO finds the best location and movement for the tUAVs at a given observation of entire network of rUAVs.

Reinforcement learning has been widely used in UAV related research recently. This includes a range of application from military threat avoidance [48] and obstacle avoidance [49] to trajectory optimization for improving services in wireless communications [50]. In our previous work [35], we used Q-Learning which is not scalable for large observation space and multi-agent systems. In our current work, we utilize DRL to solve the scalability issue where deep neural networks are used to improve reinforcement learning. Among several DRL methods such as deep Q-network (DQN), deep deterministic policy gradient (DDPG), PPO and Twin-Delayed Deep Deterministic Policy Gradient Agents (TD3), we found PPO to perform the best in terms of faster learning, relatively little hyperparameter tuning and simplicity [51]. Hence, we employ PPO with discrete observation and action space. PPO also allows us to use fine-grained discrete observation value. In contrast, discretizing observation values can be an implementation issue in Q-Learning as it increases the Q-table size sharply. PPO components in our solution are defined as the following:

- **Agent** (tUAV) observes the current state and takes actions. There are multiple agents in our scenario. To keep the model simple, we implemented multiple tUAV system as a single agent PPO with multiple actions.
- **State** ( $S$ ) is defined based on the observed information of rUAVs and the current location of tUAV. Thus, we define the system state as  $S = \{L_c, L_h, B_h\}$  where  $L_c$  is the location of tUAVs,  $L_h = [L_{h_1}, L_{h_2}, \dots, L_{h_z}]$  is a vector that denotes the locations of rUAV1 to rUAVz and  $B_h = [B_{h_1}, B_{h_2}, \dots, B_{h_z}]$  is a vector that denotes their battery levels.
- **Action** ( $a$ ) is defined as flying to hovering above certain rUAVs. Hence, the number of possible actions is equal to number of rUAVs. The PPO algorithm implements a function approximator  $\mu(S)$  that takes state  $S$  and returns the probabilities of taking each action in the action space.
- **Revenue** ( $R$ ) is the combination of rewards and penalties after taking action  $a$  at state  $S$  and moving to state  $S'$ . It returns a reward for the energy that all rUAVs receive from tUAV and/or applies a penalty if an rUAV has to move to a terrestrial charging station due to low battery.  $R$  is formulated as:

$$R(S, a, S') = w_1 \int_T P_r dt + w_2 N_o + w_3 B_l + w_4 B_f + w_5 Q \quad (8)$$

where  $w_1, \dots, w_5$  are adjusting weights,  $P_r$  is the total harvested power by rUAVs noted in (5),  $T$  is the time step and  $N_o$  is the number of out of charge rUAVs which should be replaced and resulted in service interruption.  $B_l$  represents the low battery thresholds of rUAVs and is defined as

$$B_{low} = \sum_{k=1}^Z \bar{B}_k, \quad \text{while} \quad (9)$$

$$\bar{B}_k = \begin{cases} 0.05B_{max} - B_{h_k} & \text{if } B_{h_k} \leq 0.05B_{max} \\ 0 & \text{otherwise} \end{cases},$$

where  $B_{max}$  is the battery capacity of rUAV.  $B_f$  denotes the full battery if the battery is more than 97% charged. Finally,  $Q$  indicates the conflict between the tUAVs if their

distance from each other is less than a threshold. This can force them to not charge the same rUAV in the same time and also avoid a crash. The second PPO function approximator is the critic  $V(S)$  that takes observation  $S$  and returns the expectation of the discounted long-term reward [46,47].

In the above model, each agent (tUAV) needs to observe all rUAVs' geographical locations and their remaining battery levels. We assume that the rUAVs remain in the same geo-cell in our considered area; therefore, only their battery status needs to be sent to the tUAVs at each time step. Hence, our tUAVs and rUAVs must have a light periodic signaling to exchange information.

Considering the discussed Reinforcement Learning components, we follow Algorithm 1 to obtain an optimal flying trajectory (i.e., movement decisions) of the tUAVs and recharging mechanism that maximizes the overall flying duration of all rUAVs. In this algorithm, each tUAV receives updated information of the rUAVs at each time step. The observation includes the tUAVs current location indicating the current state. The agent makes a decision on movement based on the actor output. Recharging is considered only when the tUAV arrives to hover above the chosen rUAV because recharging is assumed inefficient when the tUAV is flying. The PPO details are not presented in Algorithm 1 as it can be found in [46,47]. The algorithm can be executed centrally in a ground control station or by the tUAVs individually.

---

**Algorithm 1** tUAV Trajectory Algorithm.

---

```

Initialize Actor  $\mu(S)$  and Critic  $V(S)$  with random values
Observe rUAVs' locations
REPEAT
  Observe rUAVs Battery
  Current state = (tUAV's location, Observation)
  Update tUAV's location by taking an action for current state  $\mu(S)$ 
  Calculate Revenue of the tUAV's last movement
  If there is enough experiences, update Actor  $\mu(S)$  and Critic  $V(S)$ 
CONTINUE

```

---

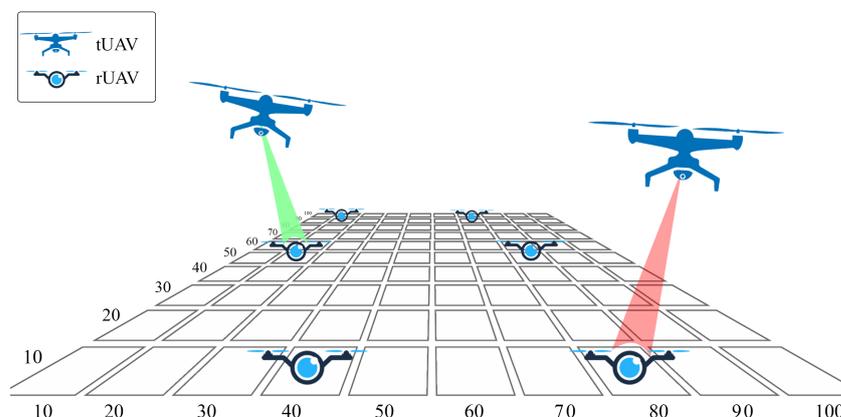
#### 4. Performance Evaluation

In this section, we first describe the simulation set-up including the baseline algorithms that we compare the PPO's performance against. We then present and discuss key results of this research.

##### 4.1. Simulation Setup

In our scenarios, we consider six rUAVs and two tUAVs, located in an environment modeled as a  $100 \times 100$  grid (Figure 2). To simplify our simulation design, we assume all rUAVs can be located only at the center of cells as illustrated in Figure 2. Each tUAV sends the recharging beam toward the target rUAV that is selected by the algorithm. We selected an arbitrary frequency of 25–27 GHz which can be adjusted as per the spectrum regulations in the region. Note that increasing the frequency increases free space path loss but more antennae can be installed in the same antenna size since the MIMO proper inter-element space is related to wavelength. For example, the wavelength of frequencies below 1 GHz is very large for MIMO. Additionally, 1–7 GHz is highly saturated for current wireless communications [40]. On the other hand, high path loss in higher frequency is helpful to minimize the interference of WPT to ground stations. Hence, we propose mmWave spectrum for our conjecture. We assume a maximum power of 1Watt is transmitted at each sub-band of 10 MHz width. This is reasonable in terms of regulations as in most countries mobile devices that work in the millimeter-wave spectrum are permitted to operate in 83 dBm/100 MHz range [23]. There are 256 antenna elements installed on each tUAV and rUAV in a uniform square array, and since the EM wavelength is about 1.2 cm, the array

can be readily fitted on a small drone. For a square array, the number of antenna elements should be a power of 2, e.g., 256. All simulation parameters are defined in Table 1.



**Figure 2.** Considered simulation scenarios showing tUAV and rUAV positions. The rUAVs are stationary in each scenario and two tUAVs periodically change position to improve the energy transfer efficiency where the initial position of tUAV is randomized for each episode.

**Table 1.** Simulation parameters.

Simulation Component	Value
Transmit power of each sub-band $P_s$	1 Watt
Antenna element gain $G_t, G_r$	16 dBi
Number of antenna on tUAV $m_t$	256
Number of antenna on rUAV $m_r$	256
Number of sub-bands $N$	200
Sub-band's width	10 MHz
Cell side	10 m
Charging Wave Frequency range	25–27 GHz
Learning rate	0.4
Discount factor	0.95
rUAV power consumption	$50 \pm 10$ Watt
rUAV battery capacity	30 Watt-hour (108 kJ)
Time step	30 or more s
Revenue adjusting weights $(w_1, w_2, w_3, w_4, w_5)$	0.001, $-10,000$ , $-0.0001$ , $-0.00003$ , $-10,000$

In order to evaluate our algorithm's performance, we used the MATLAB Reinforcement Learning Toolbox to simulate the environment and implement the PPO algorithm. Additionally, we simulated the following two benchmark schemes for the tUAVs' movement decisions:

- Traveling Salesman Problem (TSP): Each tUAV recharges a group of three rUAVs periodically and in order. The groups and orders should be selected so that the traveling times of the tUAVs are minimized. We solve the TSP using an iterative approach to find the best two groups to be served by the two tUAVs.
- Lowest Battery First (LBF): The tUAVs target to serve the rUAVs with the minimum battery level at each time step.

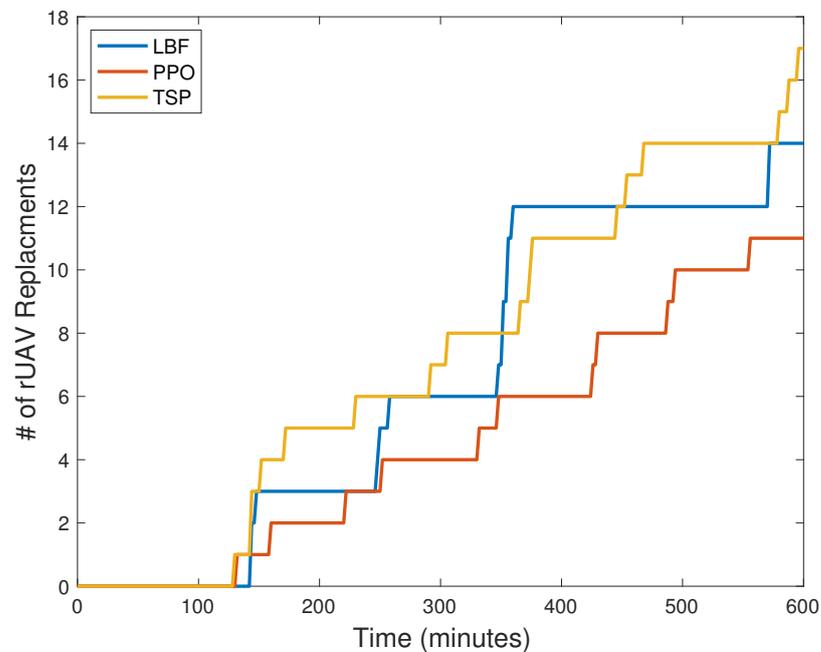
To compare the performance of the PPO and the above baseline schemes and also to show the WPT recharging effect, we counted the number of times that an rUAV battery reaches the minimum threshold and it is replaced with a full battery rUAV after few seconds. The rUAV replacement can result in a service interruption to the nodes that are served by the respective rUAV (e.g., in the Hotspot service scenario). As such, the replacements should be minimized. Additionally, we calculated the average flying time of all rUAVs. We assumed that the WPT recharging in our scenario is not enough to keep

all rUAVs in service for long time. This is because the total recharging power is less than consumed power. A period of 10 h was simulated to study the impact of the recharging.

#### 4.2. Results

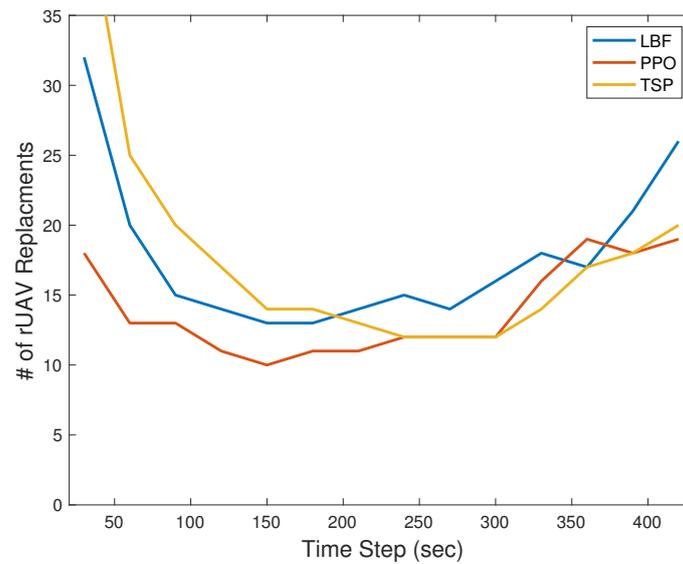
In this section, we present the simulation results based on the above system model and algorithm.

First, we ran simulations with and without the WPT capability to see the viability of our proposed model. We can evaluate both systems' performances based on the number of times the rUAVs need to be replaced due to the battery depletion. As is illustrated in Figure 3, using the MIMO based WPT-enabled tUAVs significantly improves the system performance by reducing the number of rUAV replacements from 108 (when no WPT recharging is used) to less than 20 during a 10 h simulation period. In the same simulation, we also evaluated the performance of the proposed algorithm against benchmark schemes with WPT enabled tUAVs. As it can be seen from Figure 3, PPO outperforms the benchmark schemes where in 10 h duration, the number of rUAV replacements is only 11 in comparison with 14 and 17 for the LBF and TSP schemes. We assumed 120 s time steps for tUAVs to retake decisions in this simulation.



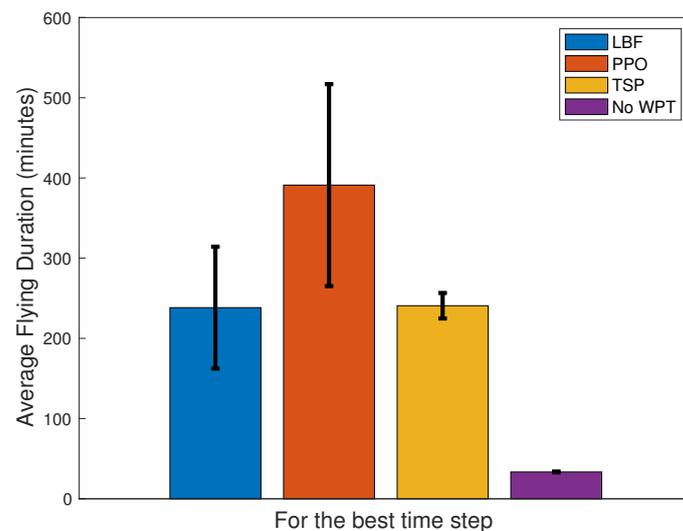
**Figure 3.** Comparison of number of rUAV replacements when an rUAV is out of charge. Note that after 600 min, 108 replacements were recorded without WPT recharging.

Second, since the tUAV's recharging ability is not used when it is traveling, increasing the time step may improve the results for benchmarks. For this purpose, we simulated the scenario with different time step durations to compare different models' performances. The result is plotted in Figure 4. As can be seen, the PPO's performance can also be improved for longer time steps of about 150–200 s. It is further observed that TSP can be as good as PPO for some time steps. However, the figure shows that the PPO performance superiority is maintained for all time step values despite the fact that the gap is reduced in 150–200 s. To conclude, the best overall performance is recorded for a time step of 150 s by PPO when only 10 replacements are recorded over 10 h.



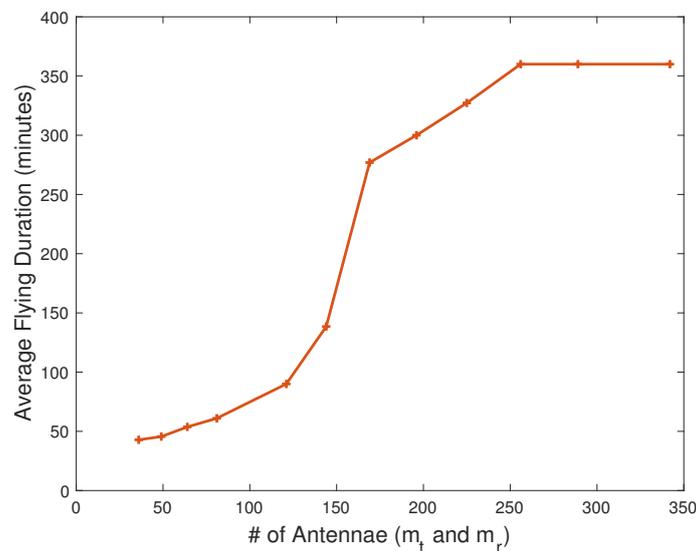
**Figure 4.** Comparison of different tUAV movement schemes used with the WPT for different time step values. Each tUAV takes the next action to update the location at the end of each time step. The lowest replacement is achieved by PPO when time step is 150 s.

Furthermore, we have recorded the flight duration of all rUAVs in our simulation. Figure 5 demonstrates their average flying times. We have used the best time step for each scheme based on Figure 4. As is shown, while it is only 33 min without employing the WPT recharging mechanism, the proposed PPO based WPT can increase rUAVs’ flight duration up to 390 min for the studied scenario. Moreover, it can be observed that the low complexity schemes of LBF and TSP can achieve an approximately 240 min flying duration, which is significant. This result demonstrates the merit of our proposed MIMO-WPT based UAVs recharging architecture, irrespective of the specific movement strategy of the tUAVs. However, the PPO’s performance achieved notable gains of about 60% higher than the benchmark movement schemes. Additionally, a confidence interval of 95% is shown in Figure 5 that shows energy distribution fairness among rUAVs. Among the three schemes of WPT, TSP is distributing energy more equally.



**Figure 5.** Average flying time of rUAVs for different time step values. In the absence of WPT recharging, the average flying time is 33 min.

On the other hand, we present the average flight duration of rUAVs in Figure 6 for different numbers of antennas in which we consider the same value for  $m_r$  and  $m_t$ . Clearly, increasing the number of antenna elements can increase the beamforming gain and improve the WPT efficiency. However, this is limited by the total transmitting power, and the WPT's efficiency (RF-RF) which cannot be more than 100%. Note that we assumed a maximum of 90% RF-RF efficiency to address the non-ideal implementation factors such as mutual impedance between antenna elements in the antenna array [43].



**Figure 6.** Average flying time of rUAVs for different antennae numbers when PPO is used.

## 5. Conclusions and Future Work

We studied the concept of using dedicated, flying chargers equipped with MIMO antenna for in situ recharging of UAVs' batteries using wireless power transfer. We formulated the movement decision of the aerial chargers to recharge the UAVs as a multi-agent optimization problem using the Proximal Policy Optimization (PPO) to optimize the energy transfer gain and enhance the UAVs' flying times. Using simulation studies, we demonstrated that the MIMO-WPT provided a tenfold increase in the flight time for the deployed system compared to no wireless recharging of the UAVs. The maximum gain was achieved when PPO was employed to place and move wireless energy sources intelligently. Although we have extracted simulation parameters and assumptions from practical works, implementation challenges may affect the gain of MIMO-WPT UAV recharging.

Although we simulated a scenario of Hotspot UAVs that hover above fixed locations, it can be generalized for all applications where the power receiver UAVs are hovering above a certain location. Future work could consider scenarios with mobility and dynamic positioning of the power-receiver UAVs, and the use of hybrid power sources at the flying chargers. Additionally, the CSI measurement at the tUAVs is not a challenge for the rotor-based rUAVs, which is what we have used in our work. However, for the winged rUAVs, this is not the case since they cannot stand still. So, in practical implementations of our system with the winged rUAVs, the CSI acquisition and dynamic beamforming will be challenging. In future work, fast beam switching technology in the face of changing CSI using ML or codebook-based beamforming can be investigated.

**Author Contributions:** Conceptualization, S.S.K.; visualization, A.B.; investigation, S.A.H.; methodology, S.A.H.; Software, A.B. and S.A.H.; writing—original draft preparation, S.A.H., A.B. and J.H.; writing—review and editing, S.S.K.; Conceptualization, J.H.; Funding acquisition, J.H.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Central Queensland University Research Grant RSH5137.

**Acknowledgments:** The authors acknowledge the Central Queensland University high performance computing resources (<https://www.cqu.edu.au/eresearch/high-performance-computing> accessed on 12 July 2021) made available for conducting the research reported in this paper. The authors also acknowledge infrastructure support from UNSW Institute for Cyber Security.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Silver, B.; Mazur, M.; Wisniewski, A.; Babicz, A. Welcome to the Era of Drone-Powered Solutions: A Valuable Source of New Revenue Streams for Telecoms Operators. 2017. Available online: <https://www.pwc.com/gx/en/communications/pdf/communications-review-july-2017.pdf> (accessed on 10 March 2020).
2. Fotouhi, A.; Ding, M.; Hassan, M. Flying Drone Base Stations for Macro Hotspots. *IEEE Access* **2018**, *6*, 19530–19539. [[CrossRef](#)]
3. Fotouhi, A.; Qiang, H.; Ding, M.; Hassan, M.; Giordano, L.G.; Garcia-Rodriguez, A.; Yuan, J. Survey on UAV Cellular Communications: Practical Aspects, Standardization Advancements, Regulation, and Security Challenges. *IEEE Commun. Surv. Tutorials* **2019**, *21*, 3417–3442. [[CrossRef](#)]
4. Zeng, Y.; Wu, Q.; Zhang, R. Accessing From the Sky: A Tutorial on UAV Communications for 5G and Beyond. *Proceedings of the IEEE* **2019**, *107*, 2327–2375. [[CrossRef](#)]
5. Zeng, Y.; Lyu, J.; Zhang, R. Cellular-Connected UAV: Potential, Challenges, and Promising Technologies. *IEEE Wirel. Commun.* **2019**, *26*, 120–127. [[CrossRef](#)]
6. Li, B.; Fei, Z.; Zhang, Y. UAV Communications for 5G and Beyond: Recent Advances and Future Trends. *IEEE Internet Things J.* **2019**, *6*, 2241–2263. [[CrossRef](#)]
7. Tran, D.H.; Vu, T.; Chatzinotas, S.; Shahbazpanahi, S.; Ottersten, B. Coarse Trajectory Design for Energy Minimization in UAV-Enabled Wireless Communications with Latency Constraints. *IEEE Trans. Veh. Technol.* **2020**, *69*, 9483–9496. [[CrossRef](#)]
8. Salehi, S.; Bokani, A.; Hassan, J.; Kanhere, S.S. AETD: An Application Aware, Energy Efficient Trajectory Design for Flying Base Stations. In Proceedings of the 2019 IEEE 14th Malaysia International Conference on Communication (MICC), Selangor, Malaysia, 2–4 December 2019.
9. Li, K.; Ni, W.; Wang, X.; Liu, R.P.; Kanhere, S.S.; Jha, S. Energy-efficient cooperative relaying for unmanned aerial vehicles. *IEEE Trans. Mob. Comput.* **2015**, *15*, 1377–1386. [[CrossRef](#)]
10. Abdulla, A.E.; Fadlullah, Z.M.; Nishiyama, H.; Kato, N.; Ono, F.; Miura, R. An optimal data collection technique for improved utility in UAS-aided networks. In Proceedings of the IEEE INFOCOM 2014—IEEE Conference on Computer Communications, Toronto, ON, Canada, 27 April–2 May 2014; pp. 736–744.
11. Zhan, C.; Zeng, Y.; Zhang, R. Energy-efficient data collection in UAV enabled wireless sensor network. *IEEE Wirel. Commun. Lett.* **2017**, *7*, 328–331. [[CrossRef](#)]
12. Abdulla, A.E.; Fadlullah, Z.M.; Nishiyama, H.; Kato, N.; Ono, F.; Miura, R. Toward fair maximization of energy efficiency in multiple UAS-aided networks: A game-theoretic methodology. *IEEE Trans. Wirel. Commun.* **2014**, *14*, 305–316. [[CrossRef](#)]
13. Morton, S.; D’Sa, R.; Papanikolopoulos, N. Solar powered UAV: Design and experiments. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 2460–2466. [[CrossRef](#)]
14. Zeng, Y.; Clerckx, B.; Zhang, R. Communications and Signals Design for Wireless Power Transmission. *IEEE Trans. Commun.* **2017**, *65*, 2264–2290. [[CrossRef](#)]
15. Huang, J.; Zhou, Y.; Ning, Z.; Gharavi, H. Wireless Power Transfer and Energy Harvesting: Current Status and Future Prospects. *IEEE Wirel. Commun.* **2019**, *26*, 163–169. [[CrossRef](#)]
16. Brown, W.C. Experiments involving a microwave beam to power and position a helicopter. *IEEE Trans. Aerosp. Electron. Syst.* **1969**, *5*, 692–702. [[CrossRef](#)]
17. Shinohara, N. Beam control technologies with a high-efficiency phased array for microwave power transmission in Japan. *Proc. IEEE* **2013**, *101*, 1448–1463. [[CrossRef](#)]
18. Strassner, B.; Chang, K. Microwave power transmission: Historical milestones and system components. *Proc. IEEE* **2013**, *101*, 1379–1396. [[CrossRef](#)]
19. Jull, G.W.; Lillemark, A.; Turner, R. SHARP (stationary high altitude relay platform) telecommunications missions and systems. In Proceedings of the GLOBECOM’85-Global Telecommunications Conference, New Orleans, LA, USA, 2–5 December 1985; Volume 2, pp. 955–959.
20. Hua, M.; Li, C.; Huang, Y.; Yang, L. Throughput Maximization for UAV-enabled Wireless Power Transfer in Relaying System. In Proceedings of the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 11–13 October, 2017; pp. 1–5. [[CrossRef](#)]
21. Ansari, N.; Wu, D.; Sun, X. FSO as backhaul and energizer for drone-assisted mobile access networks. *ICT Express* **2020**, *6*, 139–144. [[CrossRef](#)]
22. Nangia, R.K. ‘Greener’ civil aviation using air-to-air refuelling—Relating aircraft design efficiency and tanker offload efficiency. *Aeronaut. J. (1968)* **2007**, *111*, 589–592. [[CrossRef](#)]

23. Federal Communications Commission. FCC-Use of Spectrum Bands Above 24 GHz For Mobile Radio Services. 2016. Available online: [https://apps.fcc.gov/edocs\\_public/attachmatch/FCC-16-89A1.pdf](https://apps.fcc.gov/edocs_public/attachmatch/FCC-16-89A1.pdf) (accessed on 7 July 2021).
24. Hoseini, S.A.; Bokani, A.; Hassan, J.; Salehi, S.; Kanhere, S.S. Energy and Service-Priority aware Trajectory Design for UAV-BSs using Double Q-Learning. In Proceedings of the 2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2021; pp. 1–4. [CrossRef]
25. Bushnaq, O.M.; Kishk, M.A.; Celik, A.; Alouini, M.S.; Al-Naffouri, T.Y. Optimal Deployment of Tethered Drones for Maximum Cellular Coverage in User Clusters. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 2092–2108. [CrossRef]
26. Lee, D.; Zhou, J.; Lin, W.T. Autonomous battery swapping system for quadcopter. In Proceedings of the 2015 International Conference on Unmanned Aircraft Systems (ICUAS), Denver, CO, USA, 9–12 June 2015; pp. 118–124. [CrossRef]
27. Shukla, H. This Wireless Power Technology Could Change New Zealand’s Transmission System. 2020. Available online: <https://mercomindia.com/this-wireless-new-zealand-transmission-system/> (accessed on 23 June 2021).
28. Delbert, C. The Dawn of Wireless Electricity Is Finally Upon Us. Here’s How New Zealand Will Do It. 2021. Available online: <https://www.popularmechanics.com/science/a33522699/wireless-electricity-new-zealand/> (accessed on 23 June 2021).
29. Boyle, A. PowerLight Is Hitting Its Targets with a Power Beaming System That Uses Lasers. 2021. Available online: <https://www.geekwire.com/2021/powerlight-hitting-targets-power-beaming-system-uses-laser-light/> (accessed on 23 June 2021).
30. Bennett, T. TransGrid Deploys Drones to Perform Power Line Work. 2020. Available online: <https://www.itnews.com.au/news/transgrid-deploys-drones-to-perform-power-line-work-559095> (accessed on 23 June 2021).
31. Wing. Available online: [https://wing.com/en\\_au/australia/canberra/](https://wing.com/en_au/australia/canberra/) (accessed on 23 June 2021).
32. Metz, C. Police Drones Are Starting to Think for Themselves. 2020. Available online: <https://www.nytimes.com/2020/12/05/technology/police-drones.html> (accessed on 23 June 2021).
33. Banga, B. Global Autonomous Drone Wireless Charging and Infrastructure Market to Reach \$249.3 Million by 2024. 2020. Available online: <https://www.bloomberg.com/press-releases/2020-01-07/global-autonomous-drone-wireless-charging-and-infrastructure-market-to-reach-249-3-million-by-2024> (accessed on 23 June 2021).
34. Hassan, J.; Bokani, A.; Kanhere, S.S. Recharging of Flying Base Stations using Airborne RF Energy Sources. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW), Marrakech, Morocco, 15–18 April 2019; pp. 1–6. [CrossRef]
35. Hoseini, S.A.; Hassan, J.; Bokani, A.; Kanhere, S.S. Trajectory Optimization of Flying Energy Sources using Q-Learning to Recharge Hotspot UAVs. In Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 6–9 July 2020; pp. 683–688. [CrossRef]
36. Xu, J.; Bi, S.; Zhang, R. Multiuser MIMO Wireless Energy Transfer With Coexisting Opportunistic Communication. *IEEE Wirel. Commun. Lett.* **2015**, *4*, 273–276. [CrossRef]
37. Xu, J.; Zhang, R. A General Design Framework for MIMO Wireless Energy Transfer with Limited Feedback. *IEEE Trans. Signal Process.* **2016**, *64*, 2475–2488. [CrossRef]
38. Wang, Y.; Liu, A.; Xu, K.; Xia, X. Energy and Information Beamforming in Airborne Massive MIMO System for Wireless Powered Communications. *Sensors* **2018**, *18*, 3540. [CrossRef]
39. Tse, D.; Viswanath, P. *Chapter 07: MIMO I: Spatial Multiplexing and Channel Modeling; Fundamentals of Wireless Communication*; Cambridge University Press: Cambridge, UK, 2005; pp. 290–331.
40. Hoseini, S.A.; Ding, M.; Hassan, M.; Chen, Y. Analyzing the Impact of Molecular Re-Radiation on the MIMO Capacity in High-Frequency Bands. *IEEE Trans. Veh. Technol.* **2020**, *69*, 15458–15471. [CrossRef]
41. Hamdy, M.N. Beamformers Explained. 2020. Available online: <https://www.commscope.com/globalassets/digizuite/542044-Beamformer-Explained-WP-114491-EN.pdf> (accessed on 7 July 2021).
42. Agrawal, T.; Srivastava, S. Two element MIMO antenna using Substrate Integrated Waveguide (SIW) horn. In Proceedings of the 2016 International Conference on Signal Processing and Communication (ICSC), Noida, India, 26–28 December 2016; pp. 508–511. [CrossRef]
43. Aoki, T.; Yuan, Q.; Quang-Thang, D.; Okada, M.; Hsu, H.M. Maximum transfer efficiency of MIMO-WPT system. In Proceedings of the 2018 IEEE Wireless Power Transfer Conference (WPTC), Montreal, QC, Canada, 3–7 June 2018; pp. 1–3.
44. Carvalho, A.; Carvalho, N.; Pinho, P.; Goncalves, R. Wireless power transmission and its applications for powering Drone. In Proceedings of the 8th Congress of the Portuguese Committee of URSI, Lisbon, Portugal, 28 November 2014.
45. Brown, W.C. The history of wireless power transmission. *Sol. Energy* **1996**, *56*, 3–21. [CrossRef]
46. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
47. MathWorks. Proximal Policy Optimization Agents. Available online: <https://www.mathworks.com/help/reinforcement-learning/ug/ppo-agents.html> (accessed on 30 May 2021).
48. Yan, C.; Xiang, X.; Wang, C. Towards Real-Time Path Planning through Deep Reinforcement Learning for a UAV in Dynamic Environments. *J. Intell. Robot. Syst.* **2019**, *98*, 297–309. [CrossRef]
49. Yijing, Z.; Zheng, Z.; Xiaoyi, Z.; Yang, L. Q learning algorithm based UAV path learning and obstacle avoidance approach. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 3397–3402.

- 
50. Challita, U.; Saad, W.; Bettstetter, C. Deep reinforcement learning for interference-aware path planning of cellular-connected UAVs. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–7.
  51. Schulman, J.; Wolski, P.; Dhariwal, P.; Radford, D.; Fiedel, D. Proximal Policy Optimization. 2017. Available online: <https://openai.com/blog/openai-baselines-ppo/> (accessed on 15 August 2021).