sensors

MDPI

Article

Deep Q-Learning-Based Transmission Power Control of a High Altitude Platform Station with Spectrum Sharing

Seongjun Jo [1], Wooyeol Yang [1], Haing Kun Choi [2], Eonsu Noh [3], Han-Shin Jo [1,*] and Jaedon Park [3,*]

[1] Department of Electronic Engineering, Hanbat National University, Daejeon 34158, Korea; 30211176@edu.hanbat.ac.kr (S.J.); 30211173@edu.hanbat.ac.kr (W.Y.)
[2] TnB Radio Tech., Seoul 08504, Korea; radioeng@tnbrt.com
[3] Agency for Defense Development, Daejeon 34186, Korea; nes@add.re.kr
* Correspondence: hsjo@hanbat.ac.kr (H.-S.J.); jaedon2@add.re.kr (J.P.)

Abstract: A High Altitude Platform Station (HAPS) can facilitate high-speed data communication over wide areas using high-power line-of-sight communication; however, it can significantly interfere with existing systems. Given spectrum sharing with existing systems, the HAPS transmission power must be adjusted to satisfy the interference requirement for incumbent protection. However, excessive transmission power reduction can lead to severe degradation of the HAPS coverage. To solve this problem, we propose a multi-agent Deep Q-learning (DQL)-based transmission power control algorithm to minimize the outage probability of the HAPS downlink while satisfying the interference requirement of an interfered system. In addition, a double DQL (DDQL) is developed to prevent the potential risk of action-value overestimation from the DQL. With a proper state, reward, and training process, all agents cooperatively learn a power control policy for achieving a near-optimal solution. The proposed DQL power control algorithm performs equal or close to the optimal exhaustive search algorithm for varying positions of the interfered system. The proposed DQL and DDQL power control yields the same performance, which indicates that the actional value overestimation does not adversely affect the quality of the learned policy.

Keywords: Deep Q-learning (DQL); Double Deep Q-learning (DDQL); dynamic spectrum sharing; High Altitude Platform Station (HAPS); cellular communications; power control; interference management

Citation: Jo, S.; Yang, W.; Choi, H.K.; Noh, E.; Jo, H.-S.; Park, J. Deep Q-Learning-Based Transmission Power Control of a High Altitude Platform Station with Spectrum Sharing. *Sensors* **2022**, *22*, 1630. https://doi.org/10.3390/s22041630

Academic Editor: Margot Deruyck

Received: 12 January 2022
Accepted: 18 February 2022
Published: 19 February 2022

1. Introduction

A High Altitude Platform Station (HAPS) is a network node operating in the stratosphere at an altitude of approximately 20 km. The International Telecommunication Union (ITU) defines a HAPS in Article 1.66A as "A station on an object at an altitude of 20 to 50 km and a specified, nominal, fixed point relative to the Earth". Various studies have been performed on HAPS in recent years, and the commercial applications of HAPS have significantly increased [1]. In addition, the HAPS has potential as a significant component of wireless network architectures [2]. It is also an essential component of next-generation wireless networks, with considerable potential as a wireless access platform for future wireless communication systems [3–5].

Because the HAPS is located at high altitudes ranging from 20 to 50 km, the HAPS-to-ground propagation generally experiences lower path loss and a higher line-of-sight probability than typical ground-to-ground propagation. Thus, the HAPS can provide a high data rate for wide coverage; however, it is likely to interfere with various other terrestrial services, e.g., fixed, mobile, and radiolocation. The World Radiocommunication Conference 2019 (WRC-19) adopted a HAPS as the IMT Base Station (HIBS) in the frequency bands below 2.7 GHz previously identified for IMT by Resolution 247 [6], which addresses the potential interference of HAPS with an existing service. In such a situation,

if the existing service is not safe from HAPS interference, the two systems cannot coexist. Therefore, the HAPS transmitter is requested to reduce its transmission power to satisfy the interference–to–noise ratio (INR) requirement for protecting the receiver of the existing service. However, if the HAPS transmission power is excessively reduced, the signal–to–interference–plus–noise ratio (SINR) of the HAPS downlink decreases; thus, the outage probability may exceed the desired level. Herein, a HAPS transmission power control algorithm is proposed that aims to minimize the outage probability of the HAPS downlink while satisfying the INR requirement for protecting incumbents.

1.1. Related Works

Studies have been performed on improving the performance of HAPS. In [7], resource allocation for an Orthogonal Frequency Division Multiple Access (OFDMA)-based HAPS system that uses multicasting in the downlink to maximize the number of user terminals by maximizing the radio resources was studied. The authors of [8] proposed a wireless channel allocation algorithm for a HAPS 5G massive multiple-input multiple-output (MIMO) communication system based on reinforcement learning. Combining Q-learning and backpropagation neural networks allows the algorithm to learn intelligently for varying channel load and block conditions. In [9], a criterion for determining the minimum distance in a mobile user access system was derived, and a channel allocation approach based on predicted changes in the number of users and the call volume was proposed.

Additionally, spectrum sharing studies on HAPS have been performed. In [10], a spectrum sharing study was conducted to share a fixed service using a HAPS with other services in the 31/28-GHz band. Interference mitigation techniques were introduced, e.g., increasing the minimum operational elevation angle or improving the antenna radiation pattern to facilitate sharing with other services. In addition, the possibility of dynamic channel allocation was analyzed. In [11], sharing between a HAPS and a fixed service in the 5.8-GHz band was investigated using a coexistence methodology based on a spectrum emission mask.

In contrast to previous studies in which HAPS communication improvement and spectrum sharing were dealt with separately, in the present study, a combination of spectrum sharing with other systems and HAPS downlink coverage improvement is considered. In this regard, this study is more advanced than previous HAPS-related studies.

Deep Q-learning (DQL) is a reinforcement learning algorithm that applies deep neural networks to reinforcement learning to solve complex problems in the real world. DQL is widely used in various fields, including UAV, drone, and HAPS. In [12], the optimal UAV-BS trajectory was presented using a DQL for optimal placement of UAVs, and the author of [13] used a DQL to determine the optimal link between two UAV nodes. In [14], a DQL is used to find the optimal flight parameters for the collision-free trajectory of the UAV. In [15], two-hop communication was considered to optimize the drone base station trajectory and improve network performance, and a DQL was used to solve the joint two-hop communication scenario. In [16], a DQL was used for multiple-HAPS coordination for communications area coverage. A Double Deep Q-learning (DDQL) is an algorithm developed to prevent the overestimation of a DQL and shows better performance than the DQL in various fields [17].

1.2. Contributions

The contributions of the present study are as follows. (1) For the first time, a multiagent DQL was used to improve the HAPS outage performance and solve the problem of spectrum sharing with existing services. (2) We defined the power control optimization problem to minimize the outage probability of the HAPS downlink under the interference constraint for protecting the existing system. The state and reward for the training agent were designed to consider the objective function and constraints of the optimization problem. (3) Because the HAPS has a multicell structure, the number of power combinations increases exponentially as the number of cells (N_{cell}) and power levels increase linearly.

Thus, the optimal exhaustive search method requires an impractically long computation time to solve the multicell power optimization problem. The proposed DQL algorithm performs comparably to an optimal exhaustive search with a feasible computation time. (4) Even for varying positions of the interfered system, the proposed DQL produces a proper power control policy, maintaining stable performance. (5) Comparing the proposed DQL algorithm with the DDQL algorithm shows no performance degradation due to overestimation in the proposed DQL. The remainder of this paper is organized as follows.

Section 2 presents the system model, including the system deployment model, HAPS model, interfered system model, and path loss model. In Section 3, the downlink SINR and INR are calculated. In Section 4, a DQL-based HAPS power control algorithm is proposed. Section 5 presents the simulation results, and Section 6 concludes the paper.

2. System Model

2.1. System Deployment Model

HAPS communication networks are assumed to consist of a single HAPS, multiple ground user equipment (*UE*) devices (referred to as *UE*s hereinafter), and a ground interfered receiver. The HAPS, *UE*, and interfered receiver are distributed in the three-dimensional Cartesian coordinate system, as shown in Figure 1. The coordinates of the HAPS antenna and the interfered receiver antenna are (0, 0, h_{HAPS}) and (X, Y, h_V), respectively. The N_{UE} *UE* devices with an antenna height of h_{UE} are uniformly distributed within the circular HAPS area.

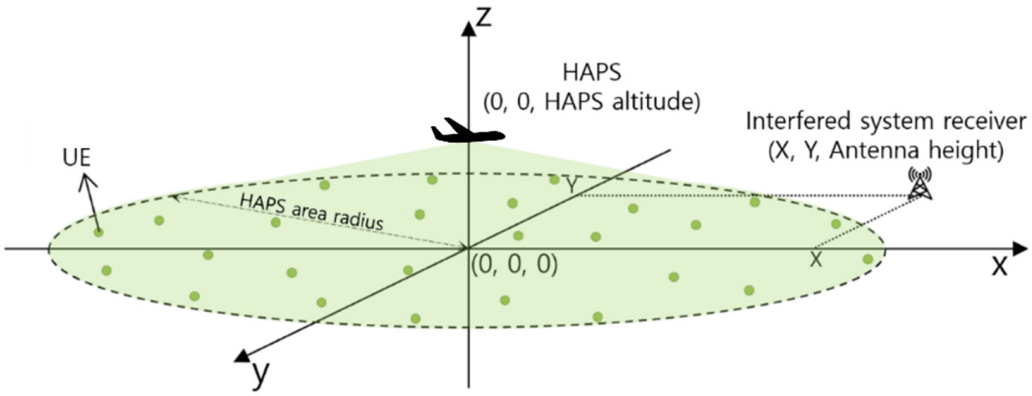


Figure 1. System deployment model.

2.2. HAPS Model

We modeled the HAPS cell deployment and system parameters with reference to the working document for a HAPS coexistence study performed in preparation for WRC-23 [18]. As shown in Figure 2, a single HAPS serves multiple cells that consist of one 1st layer cell denoted as *Cell*_1 and six 2nd layer cells denoted as *Cell*_2 to *Cell*_7. The six cells of the 2nd layer are arranged at intervals of 60° in the horizontal direction. Figure 3 presents a typical HAPS antenna design for seven-cell structures [4], where seven phased-array antennas conduct beamforming toward the ground to form seven cells, as shown in Figure 2. The 1st layer cell has an antenna tilt of 90°, i.e., perpendicular to the ground; the 2nd layer cell has an antenna tilt of 23°.

The antenna pattern of the HAPS was designed using the antenna gain formula presented in Recommendation ITU-R M.2101 [19]. The transmitting antenna gain is calculated as the sum of the gain of a single element and the beamforming gain of a multi-antenna array. The single element antenna gain is determined by the azimuth angle (ϕ) and the elevation angle (θ) between the transmitter and receiver and is calculated as follows:

$$A_E(\phi, \theta) = G_{E,max} - min\{-[A_{E,H}(\phi) + A_{E,v}(\theta)], A_m\}\ , \quad (1)$$

where $G_{E,max}$ represents the maximum antenna gain of a single element, $A_{E,H}(\phi)$ represents the horizontal radiation pattern calculated using Equation (2), and $A_{E,v}(\theta)$ represents the vertical radiation pattern calculated using Equation (3).

$$A_{E,H}(\phi) = -min\left[12\left(\frac{\phi}{\phi_{3dB}}\right)^2, A_m\right] \quad (2)$$

Here, ϕ_{3dB} represents the horizontal 3 dB beamwidth of a single element, and A_m represents the front-to-back ratio.

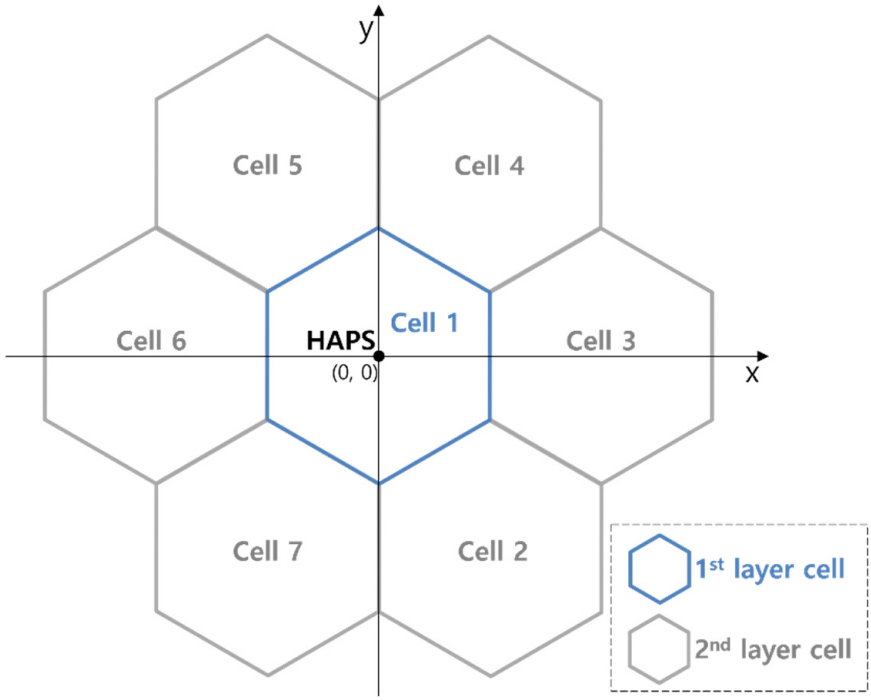


Figure 2. HAPS seven-cell layout.

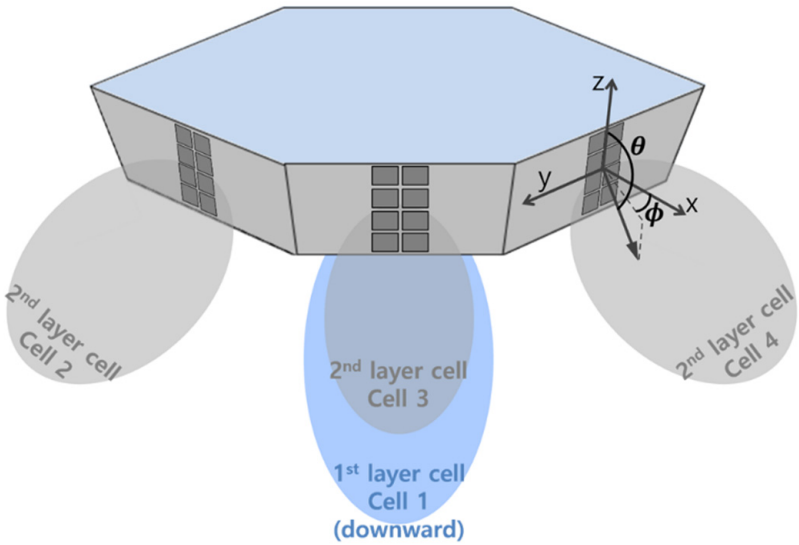


Figure 3. Typical antenna structure for multi-cell HAPS communication.

$$A_{E,V}(\theta) = -min\left[12\left(\frac{\theta - 90}{\theta_{3\text{dB}}}\right)^2, SLA_v\right] \tag{3}$$

Here, $\theta_{3\text{dB}}$ represents the vertical 3 dB bandwidth of a single element, and SLA_v represents the front-to-back ratio.

The transmitting antenna gain of the HAPS is calculated using the antenna arrangement and spacing, as well as the target beamforming direction. The gain for beam i is calculated as follows:

$$A_{A,Beami}(\theta,\phi) = A_E(\theta,\phi) + 10\log_{10}\left(\left|\sum_{m=1}^{N_H}\sum_{n=1}^{N_V} w_{i,n,m}\cdot v_{n,m}\right|^2\right), \tag{4}$$

where N_H and N_V represent the number of antennas in the horizontal and vertical directions, respectively. $v_{n,m}$ is the superposition vector that overlaps the beams of the antenna elements, which is calculated using Equation (5), and $w_{i,n,m}$ is the weight that directs the antenna element in the beamforming direction, which is calculated using Equation (6).

$$\begin{aligned} &n = 1,2,\ldots N_V; m = 1,2,\ldots N_H \\ &v_{n,m} = exp\left(\sqrt{-1}\cdot 2\pi\left((n-1)\cdot\tfrac{d_V}{\lambda}\cdot cos(\theta) + (m-1)\cdot\tfrac{d_H}{\lambda}\cdot sin(\theta)\cdot sin(\phi)\right)\right) \end{aligned} \tag{5}$$

Here, d_H and d_V represent the intervals between the horizontal and vertical antenna arrays, respectively, and λ represents the wavelength.

$$\begin{aligned} w_{i,n,m} = \tfrac{1}{\sqrt{N_H N_V}} \quad & exp\left(\sqrt{-1}\right. \\ & \cdot 2\pi\left((n-1)\cdot\tfrac{d_V}{\lambda}\cdot sin(\theta_{i,etilt}) - (m-1)\cdot\tfrac{d_H}{\lambda}\cdot cos(\theta_{i,etilt})\right. \\ & \left.\left.\cdot sin(\phi_{i,escan})\right)\right) \end{aligned} \tag{6}$$

Here, $\phi_{i,escan}$ and $\theta_{i,etilt}$ represent the ϕ and θ of the main beam direction, respectively.

The 1st layer cell of the HAPS uses a 2 × 2 antenna array, and the 2nd layer cell uses a 4 × 2 antenna array. Figure 4 shows the antenna pattern of the 1st layer cell, and Figure 5 shows the antenna pattern of the 2nd layer cell.

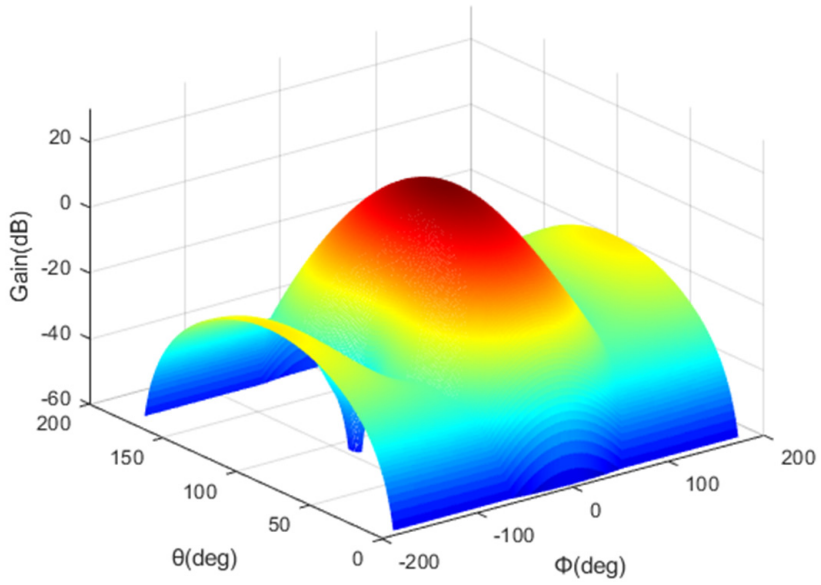


Figure 4. 1st layer cell antenna pattern.

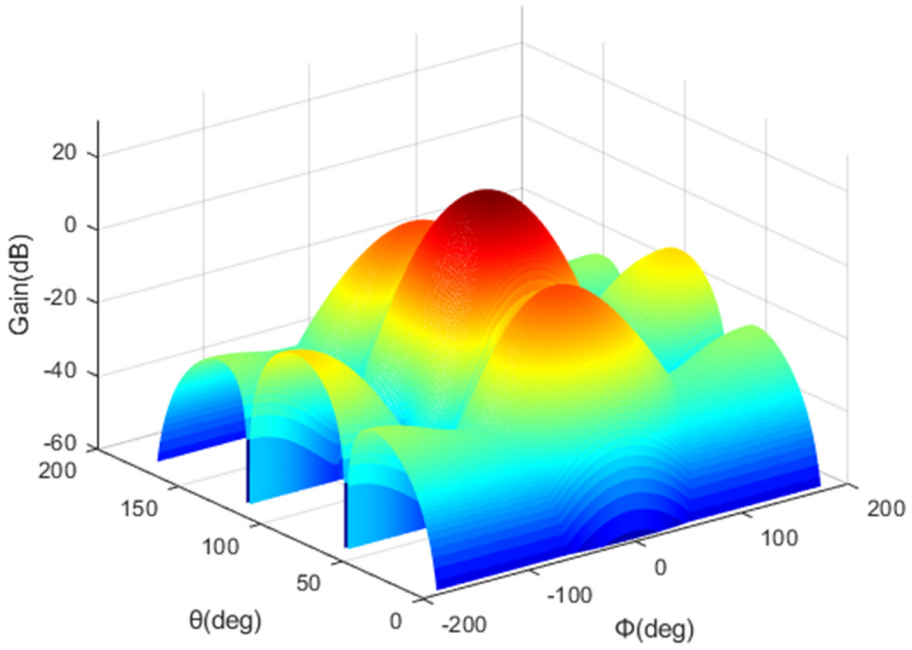


Figure 5. 2nd layer cell antenna pattern.

2.3. Interfered System Model

Various interfered systems, e.g., fixed, mobile, and radiolocation services, can be considered for the interference scenario involving a HAPS. We adopted a ground IMT base station (BS) for the interfered system, referring to the potential interference scenario [6]. The antenna pattern of the interfered system was applied by referring to Recommendation ITU-R F.1336 [20]. The receiving antenna gain is calculated as follows:

$$G(\phi, \theta) = G_0 + G_{hr}(x_h) + R \cdot G_{vr}(x_v), \tag{7}$$

where G_0 represents the maximum gain in the azimuth plane; $G_{hr}(x_h)$ represents the relative reference antenna gain in the azimuth plane in the normalized direction of $(x_h, 0)$, which is calculated using Equation (8); and $G_{vr}(x_v)$ represents the relative reference antenna gain in the elevation plane in the normalized direction of $(0, x_v)$, which is calculated using Equation (9). R represents the horizontal gain compression ratio when the azimuth angle is shifted from 0° to ϕ, which is calculated using Equation (10).

$$\begin{array}{ll} G_{hr}(x_h) = -12x_h^2 & for\ x_h \le 0.5 \\ G_{hr}(x_h) = -12x_h^{(2-k_h)} - \lambda_{\mathrm{kh}} & for\ 0.5 < x_h \\ G_{hr}(x_h) \ge G_{180} & \end{array} \tag{8}$$

$$\begin{array}{ll} G_{\mathrm{vr}}(x_v) = -12x_v^2 & for\ x_v < x_k \\ G_{\mathrm{vr}}(x_v) = -15 + 10\log(x_v^{-1.5} + k_v) & for\ x_k \le x_v < 4 \\ G_{\mathrm{vr}}(x_v) = -\lambda_{\mathrm{kv}} - 3 - C\log(x_v) & for\ 4 \le x_v < 90/\theta_3 \\ G_{\mathrm{vr}}(x_v) = G_{180} & for\ x_v \ge 90/\theta_3 \end{array} \tag{9}$$

$$R = \frac{G_{hr}(x_h) - G_{hr}(180°/\phi_3)}{G_{hr}(0) - G_{hr}(180°/\phi_3)} \tag{10}$$

Here, x_h and λ_{kh} are given by Equations (11) and (12), respectively; ϕ_3 represents the 3 dB beamwidth in the azimuth plane; and k_h is an azimuth pattern adjustment factor based on the leaked power. The relative minimum gain G_{180} was calculated using Equation (13).

$$x_h = |\phi\ |/\phi_3 \tag{11}$$

$$\lambda_{\mathrm{kh}} = 3\left(1 - 0.5^{-k_h}\right) \tag{12}$$

$$G_{180} = -15 + 10\log(1 + 8k_a) - 15\log\left(\frac{180^\circ}{\theta_3}\right) \tag{13}$$

Returning to Equation (9), x_v is given by Equation (14), and the 3-dB beamwidth in the elevation plane θ_3 is calculated using Equation (15), where G_0 represents the maximum gain in the azimuth plane. In addition, x_k is calculated using Equation (16), where k_v is an elevation pattern adjustment factor based on the leaked power. λ_{kv} was calculated using Equation (17), and the attenuation inclination factor C was calculated using Equation (18). Figure 6 shows the antenna pattern of the interfered system calculated using Equation (7), which is the pattern for a typical terrestrial BS with a broad beamwidth in the azimuth plane but a narrow beamwidth in the elevation plane.

$$x_v = |\theta\,|/\theta_3 \tag{14}$$

$$\theta_3 = 107.6 \times 10^{-0.1G_0} \tag{15}$$

$$x_k = \sqrt{1.33 - 0.33k_v} \tag{16}$$

$$\lambda_{\mathrm{kv}} = 12 - C\log(4) - 10\log\left(4^{-1.5} + k_v\right) \tag{17}$$

$$C = \frac{10\log\left(\frac{\left(\frac{180^\circ}{\theta_3}\right)^{1.5}\cdot\left(4^{-1.5}+k_v\right)}{1+8k_p}\right)}{\log\left(\frac{22.5^\circ}{\theta_3}\right)} \tag{18}$$

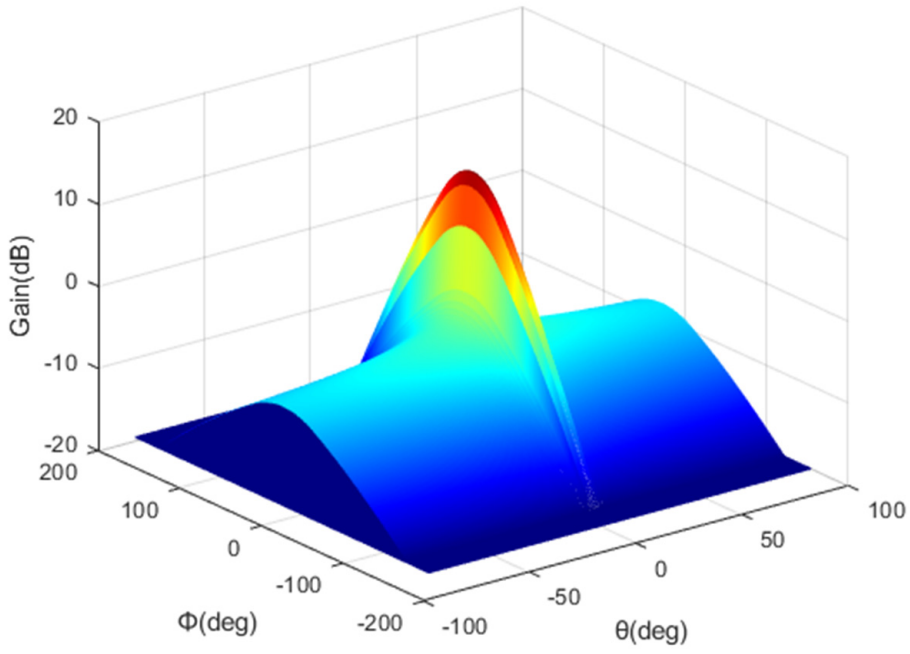


Figure 6. Interfered system antenna pattern.

2.4. Path Loss Model

The path loss model of Recommendation ITU-R P.619 [21] was applied to the working document for the HAPS coexistence study performed in preparation for WRC-23 [22]. The total path loss that occurs when the HAPS signal reaches the *UE* and the IMT BS is expressed as follows:

$$L_p = FSL + A_{xp} + A_g + A_{bs}, \tag{19}$$

where FSL represents the free-space path loss calculated using Equation (20), which occurs in a straight path from a transmitting antenna to a receiving antenna in a vacuum state, and A_{xp} is assumed to be 3 dB for depolarization attenuation. A_g represents the attenuation loss due to atmospheric gases. A_{bs} represents the resistive loss due to the spread of the antenna beam as the beam spreads attenuation. A_g and A_{bs} were calculated using the formulae in P.619.

$$FSL = 92.45 + 20\log(f \cdot d) \tag{20}$$

Here, f represents the carrier frequency (in GHz), and d represents the distance (in km) between the transmitter and receiver.

3. Calculation of Downlink SINR and INR

3.1. Calculation of Downlink SINR

The signal received by the *UE* from the HAPS transmission for the *i*th cell (*Cell_i*) is calculated as follows:

$$S_{Cell\_i} = P_{Cell\_i} + G_{Cell\_i} + G_p + G_{r,UE} - L_p - L_{ohm}, \tag{21}$$

where P_{Cell_i} represents the HAPS transmission power for *Cell_i*, G_{Cell_i} represents the transmitting antenna gain of *Cell_i*, G_p represents the polarization gain, $G_{r,UE}$ represents the receiving antenna gain, and L_{ohm} represents the ohmic loss. The *UE* receives signals from all N_{cell} cells and considers the remaining signals (except for the strongest *Cell j* signal) as interference. Equation (22) is used to calculate the signal and interference, and the receiver noise is calculated using Equation (23).

$$\begin{aligned} j &= \underset{i}{\text{argmax}} S_{Cell\_i} \\ S_{HAPS} &= S_{Cell\_j} \\ I_{HAPS,UE} &= 10\log(\sum_{\substack{i=1 \\ i \neq j}}^{N_{cell}} 10^{\frac{S_{Cell\_i}}{10}}) \end{aligned} \tag{22}$$

$$N = 10\log(k \times T \times BW) + N_f \tag{23}$$

Here, k and T represent the Boltzmann constant and noise temperature, respectively, and BW represents the channel bandwidth. N_f represents the noise figure. Finally, the downlink SINR is calculated as follows:

$$\eta = 10\log\left(\frac{10^{\frac{S_{HAPS}}{10}}}{10^{\frac{I_{HAPS,UE}}{10}} + 10^{\frac{N}{10}}}\right). \tag{24}$$

3.2. Calculation of INR

The interference power received by the interfered receiver from the HAPS transmitter servicing *Cell i* is calculated as follows:

$$I_{Cell\_i} = P_{Cell\_i} + G_{Cell\_i} + G_p + G_{r,V} - L_p - L_{ohm}, \tag{25}$$

where $G_{r,V}$ represents the antenna gain of the interfered receiver. The aggregated interference power at the interfered receiver is calculated as follows:

$$I_{HAPS,V} = 10\log\left(\sum_{i=1}^{N_{cell}} 10^{\frac{I_{Cell\_i}}{10}}\right). \tag{26}$$

Finally, after converting the aggregated interference into INR form in accordance with Equation (27) and comparing it with the protection criteria (INR_{th}) of the interfered receiver,

it is possible to check whether the interfered receiver is protected from the interference of the HAPS.

$$INR = I_{HAPS,V} - N \tag{27}$$

4. DQL-Based HAPS Transmission Power Control Algorithm

4.1. Problem Formulation

To satisfy the INR_{th} of the interfered system, the transmission power of the HAPS must be reduced. However, as the power of the HAPS is reduced, the η of the *UE* decreases, and the outage probability P_{out} increases. Thus, the objective of this study was to find a HAPS transmission power set for each cell, i.e., $\boldsymbol{P} = \{P_{Cell\_i} | i = 1, \cdots, N_{cell}\}$, that satisfies the INR_{th} of the interfered system while minimizing P_{out}. The optimization problem of the HAPS transmission power can be formulated as follows:

$$\begin{aligned} &\min_{\boldsymbol{P}} P_{out} = \frac{N_{UE,o}(\boldsymbol{P})}{N_{UE}} \\ \text{s.t.} \quad &C1: \ INR \leq INR_{th} \\ &C2: \ P_{min} \leq P_{Cell_i} \leq P_{max} \ \ \forall i \in \{1, \cdots, N_{cell}\}, \end{aligned} \tag{28}$$

where $N_{UE,o}(\boldsymbol{P})$ represents the number of UEs that do not satisfy the minimum required SINR η_o for a given HAPS transmission power set $\boldsymbol{P}$.

4.2. Proposed Algorithm

To control the HAPS transmission power, it is necessary to independently determine the power level of each cell. Accordingly, the total number of HAPS transmission power sets increases exponentially to $N_p^{N_{cell}}$ as the number of selectable powers N_p increases linearly. Although an exhaustive search algorithm can be used to find optimal solutions, this incurs excessive complexity and a long computation time. To solve this problem, we propose a DQL-based power optimization algorithm that can find a near-optimal $\boldsymbol{P}$ with low complexity. In the proposed DQL model, each agent functions as the power controller of a cell; accordingly, the number of agents is N_{cell}.

The agent—the subject of learning—learns a deep neural network called Deep Q Network (DQN) and selects an action using this network. DQL is an improved Q-learning method. Q-learning is a method for selecting the best action in a specific state through the Q-table of a state-action pair. As the state–action space grows in Q-learning, creating a Q-table and finding the best policy become highly complex. In addition, the use of Q-learning is limited because learning in the Q-table format becomes more complex when multiple agents are used. In contrast, a DQL is a promising way to solve the curse of dimensionality by approximating a Q function using a deep neural network instead of a Q-table. The proposed algorithm uses a method in which each agent learns a policy based on its observation and action while treating all other agents as part of the environment to solve the multiple-agent problem.

The basic DQL parameters (state, action, and reward) are presented below. Each agent learns the policy independently using the training data at each timestep t. The state space of the m^{th} agent comprises a set of $(N_{cell} - 1)$ interferences that the agent provides to UEs located at the centers of other cells and the agent's interference to the interfered receiver, which is expressed as

$$\boldsymbol{S}_t = \{I_v, \{I_{UE\_i} | i = 1, \cdots, N_{cell}, \ \ and \ \ i \neq m\}\}. \tag{29}$$

Two power sets configure the action space of an agent: $\boldsymbol{A}_1 = \{29, 31, 33, 35, 37\}$ and $\boldsymbol{A}_2 = \{26, 28, 30, 32, 34\}$ (unit: dBm). The agent of *Cell*_1 in the 1st layer cell selects an action from $\boldsymbol{A}_1$, and the agents of the 2nd layer cell select an action from $\boldsymbol{A}_2$. All agent actions are initialized to the minimum power value to minimize the interference to the interfered receiver at the beginning of the learning process. The reward is calculated as follows. First, because the interfered receiver must be safe from HAPS interference,

an agent receives a fixed r_t of -100 (deficient value) for $INR > INR_{th}$. In contrast, for $INR \leq INR_{th}$, an agent receives r_t computed according to the lower 5% downlink SINR of each cell $\{\hat{\eta}_i | i = 1, 2, \cdots, N_{cell}\}$ and the required SINR η_o. The reward can be expressed as

$$r_t = \begin{cases} r_{1,t} + r_{2,t} & for\ INR \leq INR_{th} \\ r_t = -100 & otherwise, \end{cases}, \tag{30}$$

where

$$\begin{array}{ll} r_{1,t} = 10 \cdot (\sum(\hat{\eta}_i - \eta_o)) & for\ \hat{\eta}_i \geq \eta_o \\ r_{2,t} = \sum(\hat{\eta}_i + \eta_o) & for\ \hat{\eta}_i < \eta_o. \end{array} \tag{31}$$

Figure 7 shows the structure of the proposed DQL-based HAPS transmission power control algorithm. Each agent learns its DQN, and one DQN consists of the main network, target network, and replay memory. The main network estimates the Q-value $Q(s, a; w)$ corresponding to the state–action pair through a deep neural network with a weight w. The main network consists of an input layer composed of seven neurons, a hidden layer consisting of 24 neurons, and an output layer consisting of five neurons. It is a fully connected network. w is updated every t in the direction that minimizes the loss function $L(w) = \mathbb{E}\left[\left(y_j - Q(s, a; w)\right)^2\right]$. The target network calculates the target value $y_j = r_j + \gamma \max_{a'} \hat{Q}(s', a'; w^-)$, where γ is the discount factor; s' and a' denotes the state and action, respectively, in the next step; and $\hat{Q}(s', a'; w^-)$ is the Q-value estimated through the target network with weight w^-. The agent's transition tuple (s_t, a_t, r_t, s_{t+1}) is piled in the replay memory, from which a minibatch (size of 512 tuples) are randomly sampled at each step. The minibatch data are used to compute the target value y_j. In a DQL, learning is stabilized, and the learning performance is improved through replay memory and a separate target network [23].

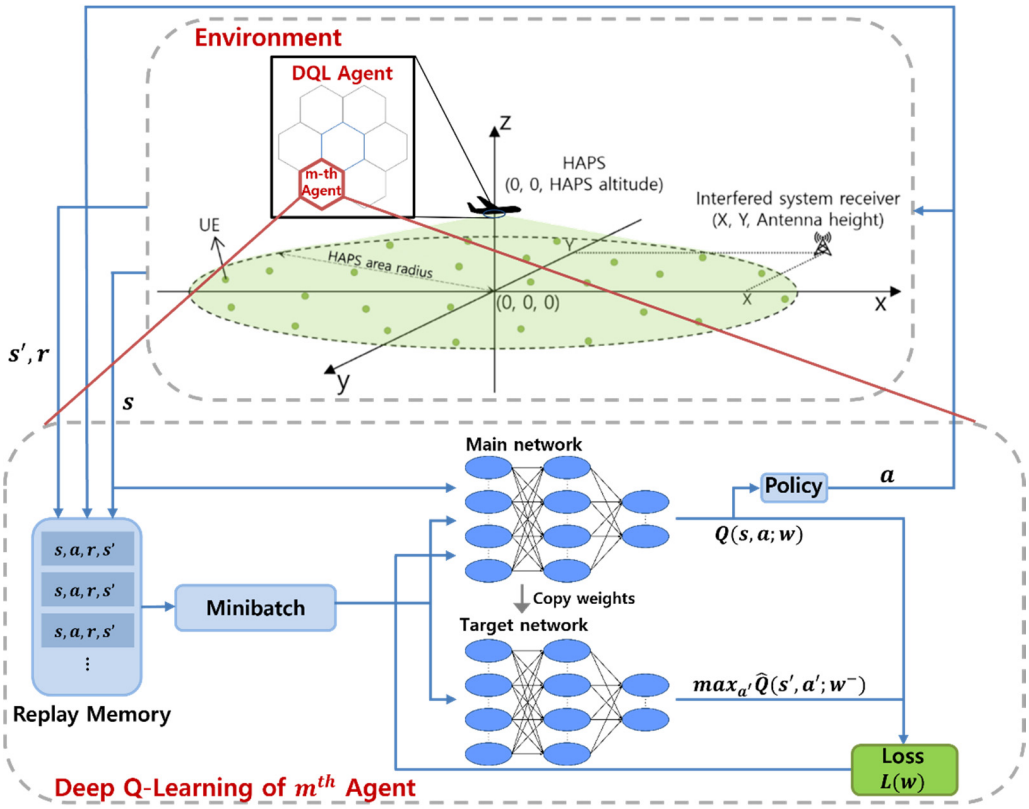


Figure 7. DQL-based HAPS power control architecture.

Algorithm 1 describes the proposed DQL-based HAPS transmission power control algorithm. For DQN training, N was set as 100,000, and the minibatch size was set as 512. M was set as 500, and T was set as 10. The Adam optimizer was used to minimize $L(w)$, and the learning rate and γ were 0.01 and 0.995, respectively. An ε-greedy policy was used to balance exploration and exploitation; ε was initially set as 1 and was reduced by 0.01 for every episode.

Algorithm 1. Training Process for the DQL-Based HAPS Power Control Algorithm

1: Initialize the replay memory $\mathcal{D}$ to capacity N
2: Initialize the Q-function with random weights w
3: Initialize the target $\hat{Q}$-function with the same weights: $w^- = w$
4: **for** episode = 1, M **do**
5: Initialize action $a_0 = \min_a A$
6: **for** timestep = 1, T **do**
7: **if** $t = 1$
8: Calculate s_t via Equations (21) and (25)
9: **end if**
10: With probability, select a random action a_t
11: Otherwise, select $a_t = \text{argmax}_a Q(s_t, a; w)$
12: Assign the selected power to the mth cell and compute INR and η
13: Observe the reward r_t and s_{t+1}
14: Store the experience in $(s_t,\ a_t,\ r_t,\ s_{t+1})$ in $\mathcal{D}$
15: Sample a random minibatch of experiences from $\mathcal{D}$
16: Set $y_j = r_j + \gamma \max_{a'} \hat{Q}(s', a'; w^-)$
17: Perform optimization via $L(w)$ and update w
18: Update the target network $\hat{Q}$ with $w^- = w$ every 4 steps
19: **end for**
20: **end for**

A DDQL is a reinforcement learning algorithm to improve performance degradation due to the overestimation of the DQL. Action-value can be overestimated by the maximization step in line 16 of Algorithm 1. Therefore, the DDQL calculates the target value as $y_j = r_j + \gamma \hat{Q}\left(s', \underset{a'}{\text{argmax}}\ Q(s', a'; w); w^-\right)$ to eliminate the maximization step. The DDQL-based HAPS power control algorithm proceeds the same way as Algorithm 1 except for calculating the target value.

5. Simulation Results

5.1. Simulation Configuration

The simulation was conducted using MATLAB for three positions of the interfered receiver, and the learning order of the agent was randomly set for each t. Subsequently, the simulation proceeded according to Algorithm 1. When all M episodes were finished, the simulation ended, and the set $\boldsymbol{P}_c$ composed of the power selected by each agent was calculated as the simulation result. Finally, the performance of the simulation was verified by comparing $\boldsymbol{P}_c$ with the optimal power set $\boldsymbol{P}^*$ obtained via an exhaustive search algorithm considering all $N_p^{N_{cell}}$ cases. The total elapsed time of the DQL and exhaustive search was about 7500 s and 21,000 s, respectively. The total elapsed time of the exhaustive search increased exponentially with the rise of N, but the DQL did not. Therefore, the computational efficiency of the DQL is more remarkable as the number of cells and power levels increase. In this simulation, performance comparison with the DDQL was additionally performed to check performance degradation due to overestimation of the DQL.

We applied the HAPS parameters and interfered system parameters, referring to the working document for the HAPS coexistence study performed in preparation for WRC-23 [18,24]. The simulation parameters of the two systems are presented in Tables 1 and 2, respectively.

Table 1. HAPS system parameters.

Parameter	Value
Center frequency (f)	2545 MHz
Channel bandwidth (BW)	20 MHz
Area radius	90 km
Altitude (h_{HAPS})	20 km
Number of cells (N_{cell})	7
Antenna pattern	Recommendation ITU-R M.2101
Element gain $(G_{E,max})$	8 dBi
Horizontal/vertical 3 dB beamwidth of single element	65° for both H/V
Antenna array configuration (Row × column)	2 × 2 elements (1st layer cell) 4 × 2 elements (2nd layer cell)
Ohmic losses (L_{ohm})	2 dB
Antenna tilt	90° (1st layer cell) 23° (2nd layer cell)
Antenna polarization	Linear/±45°
Number of distributed UEs (N_{UE})	1000
UE height	1.5 m
UE antenna gain	−3 dBi
Minimum required SINR (η_o)	−10 dB

Table 2. Interfered system (IMT BS) parameters.

Parameter	Value
Center frequency (f)	2545 MHz
Channel bandwidth (BW)	20 MHz
Noise figure $\left(N_f\right)$	5 dB
Antenna height (h_V)	20 m
Antenna tilt	10°
Antenna pattern	Recommendation ITU-R F.1336 (recommends 3.1) $k_a = 0.7$ $k_p = 0.7$ $k_h = 0.7$ $k_v = 0.3$ Horizontal 3 dB beamwidth: 65° Vertical 3 dB beamwidth is determined from the horizontal beamwidth equations in Recommendation ITU-R F.1336. Vertical beam widths of actual antennas may also be used when available.

Table 2. *Cont.*

Parameter	Value
Antenna polarization	Linear/$\pm 45°$
Maximum antenna gain (G_0)	16 dBi
Protection criteria (INR_{th})	-6 dB

5.2. Numerical Analysis

Figure 8 shows the SINR maps obtained using $\boldsymbol{P}_{max} = \{37, 34, 34, 34, 34, 34, 34\}$ and $\boldsymbol{P}_{min} = \{29, 26, 26, 26, 26, 26, 26\}$ for all cells, that is, with no power control. We considered the three positions of the interfered receiver that do not satisfy the INR_{th} of -6 dB for the use of $\boldsymbol{P}_{max}$. In addition, the three locations were designed considering the representative interference power, which can accurately reflect the operating characteristics of the proposed power control algorithm. Interfered receiver ① was located in the main beam direction for *Cell*_3 and received the highest interference from *Cell*_3. Therefore, the minimum power use of only *Cell*_3 satisfied an INR_{th} of -6 dB. Interfered receiver ② was placed on the boundary between *Cell*_3 and *Cell*_4 and thus received equal (and the strongest) interference from these two cells. Interfered receiver ③ was located in the main beam direction for *Cell*_3, as the interfered receiver. However, the minimum power use of only *Cell*_3 could not satisfy the INR_{th} of -6 dB, and at least one other cell had to use less than the maximum power.

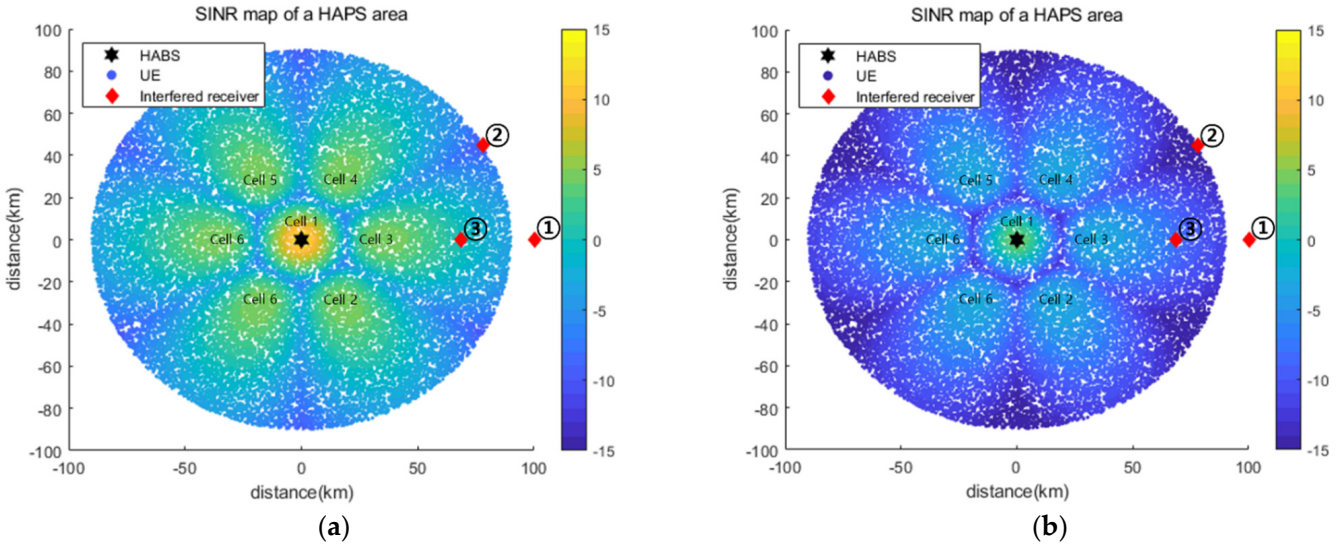


Figure 8. (**a**) SINR map for $\boldsymbol{P}_{max}$; (**b**) SINR map for $\boldsymbol{P}_{min}$.

Table 3 presents the INR and P_{out} for $\boldsymbol{P}_{max}$ and $\boldsymbol{P}_{min}$ with varying interfered receiver locations. The results confirm that the P_{out} and INR had a tradeoff relationship. The same P_{out} is shown regardless of the interference receiver position because of the absence of power control. Next, we compared the simulation results of the optimal exhaustive search and the proposed DQL-based power control algorithm for the three positions of the interfered receiver.

Table 3. INR and P_{out} for the interfered receiver locations.

Interfered Receiver	Location (km)	INR for P_{max} (dB)	INR for P_{min} (dB)	P_{out} for P_{max} (%)	P_{out} for P_{min} (%)
①	100, 0, 0.02	−3.01	−11.01	0	43.7
②	77.9, 45, 0.02	−4.08	−12.08	0	43.7
③	65.8, 0, 0.02	1.81	−6.19	0	43.7

5.2.1. Simulation Results for Interfered Receiver ①

Figure 9 shows the SINR map based on the $\boldsymbol{P}_c$ acquired using the proposed DQL-based power control algorithm for interfered receiver ①. Table 4 presents a performance comparison of the $\boldsymbol{P}^*$ values obtained via an exhaustive search and $\boldsymbol{P}_c$ and a comparison of DQL and DDQL results. As shown, $\boldsymbol{P}_c$ was equal to the optimal value $\boldsymbol{P}^*$, providing the same P_{out} and INR performance. Because the interfered receiver was located in the azimuth main beam direction of $Cell\_3$, the power of $Cell\_3$ significantly affected the interfered receiver. Even though all other cells used the maximum power, their interference was negligible. Therefore, all the cells except for $Cell\_3$ used the maximum power for minimizing P_{out}, as shown in Table 4.

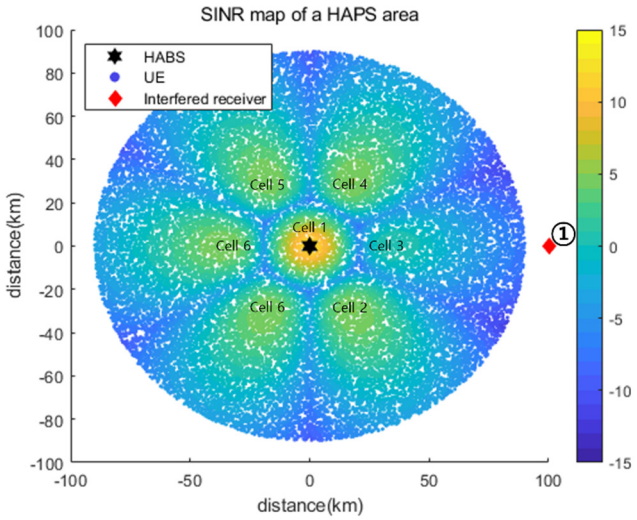


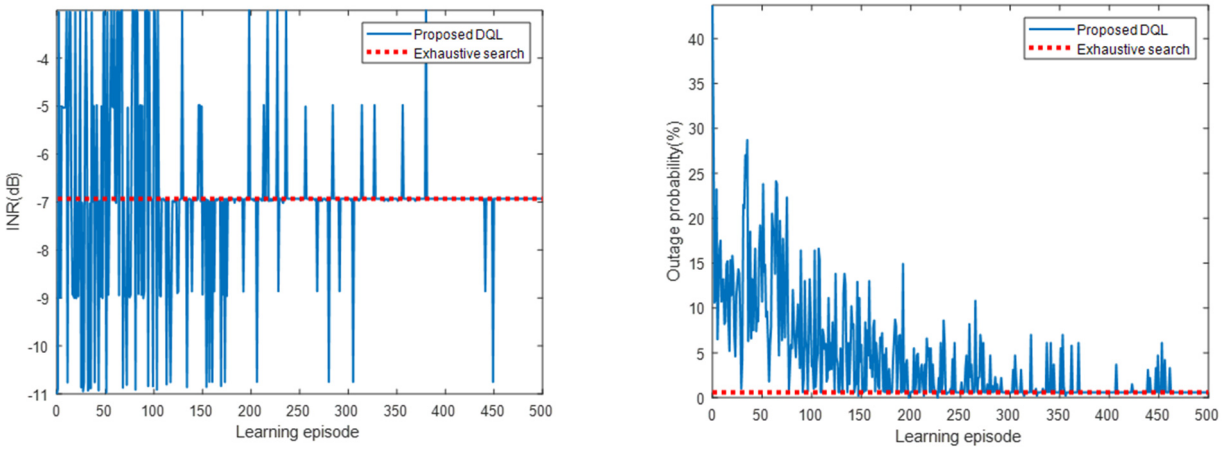
Figure 9. SINR map based on the $\boldsymbol{P}_c$ obtained using the proposed DQL-based power control algorithm for interfered receiver ①.

Table 4. Performance comparison for interfered receiver ①.

	P_{Cell_1} (dBm)	P_{Cell_2} (dBm)	P_{Cell_3} (dBm)	P_{Cell_4} (dBm)	P_{Cell_5} (dBm)	P_{Cell_6} (dBm)	P_{Cell_7} (dBm)	INR (dB)	P_{out} (%)
Optimal	37	34	30	34	34	34	34	–6.93	0.6
DQL	37	34	30	34	34	34	34	–6.93	0.6
DDQL	37	34	30	34	34	34	34	-6.93	0.6

Figure 10 presents the INR and p_{out} for each learning episode. As shown, the INR and p_{out} converged to the optimal values of the exhaustive search algorithm as the number of learning episodes increased. The INR started at −11.01 dB, which was the value for the use of $\boldsymbol{P}_{min}$, as shown in Table 3, and converged to the optimal value of −6.93 dB. Similarly, p_{out} started at 43.7% and converged to 0.6%. A large variance due to frequent exploration was observed at the beginning of the learning, but it gradually decreased and converged as the learning progressed. Figure 11 presents the cumulative and average rewards for each learning episode. As shown, the reward rapidly increased and then gradually converged at

approximately 300 episodes, indicating that the proposed DQL training process allowed the agent to learn the power control algorithm quickly and stably.



(**a**) (**b**)

Figure 10. (**a**) INR and (**b**) p_{out} for each learning episode for interfered receiver ①.

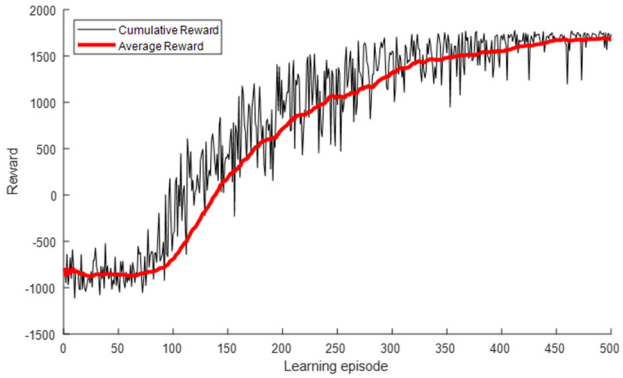


Figure 11. Reward for each learning episode for interfered receiver ①.

We compared the learning results of the DQL and DDQL. Even when the DDQL is used, the results are the same as in Table 4 and Figures 10 and 11, which shows that the overestimation of the DQL did not occur. As a result, it was confirmed that performance degradation due to overestimation did not happen, and sufficient learning is possible only with DQL.

5.2.2. Simulation Results for Interfered Receiver ②

Figure 12 shows the SINR map based on $\boldsymbol{P}_c$ acquired using the proposed DQL-based power control algorithm for interfered receiver ②. Table 5 presents a performance comparison of the $\boldsymbol{P}^*$ values obtained via an exhaustive search and $\boldsymbol{P}_c$ and a comparison of the DQL and DDQL results. As shown, $\boldsymbol{P}_c$ was equal to the optimal value $\boldsymbol{P}^*$, providing the same P_{out} and INR performance. The interfered receiver was located on the boundary between $Cell\_3$ and $Cell\_4$ and, thus, received equal (and the strongest) interference from these two cells. In addition, even though all the cells other than $Cell\_3$ and $Cell\_4$ used the maximum power, their interference was marginal. Therefore, in the optimal power control, $Cell\_3$ and $Cell\_4$ reduced the power required to satisfy the INR_{th}, whereas all the other cells used the maximum power for minimizing P_{out}, as shown in Table 5.

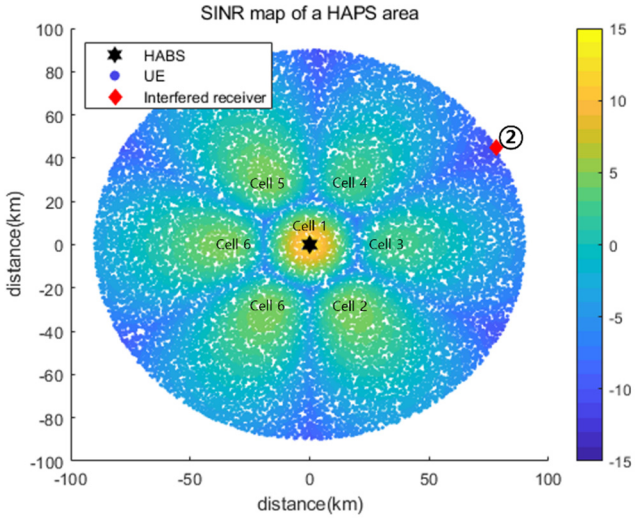


Figure 12. SINR map based on the P_c obtained using the proposed DQL-based power control algorithm for the interfered receiver ②.

Table 5. Performance comparison for interfered receiver ②.

	P_{Cell_1} (dBm)	P_{Cell_2} (dBm)	P_{Cell_3} (dBm)	P_{Cell_4} (dBm)	P_{Cell_5} (dBm)	P_{Cell_6} (dBm)	P_{Cell_7} (dBm)	*INR* (dB)	P_{out} (%)
Optimal	37	34	32	32	34	34	34	−6.08	0.2
DQL	37	34	32	32	34	34	34	−6.08	0.2
DDQL	37	34	32	32	34	34	34	−6.08	0.2

As shown in Figure 13, the INR and p_{out} converged to the optimal values of the exhaustive search algorithm. Similar to the case of receiver ①, as the learning progressed, the INR converged from −12.08 to −6.08 dB, and the p_{out} converged from 43.7% to 0.2%. Figure 14 shows that the reward gradually converged at approximately 300 episodes, indicating that the proposed DQL training process allowed the agent to quickly and stably learn the power control algorithm. We compared the learning results of the DQL and DDQL. Even when the DDQL was used, the results were the same as in Table 5 and Figures 13 and 14, verifying that the desired learning is attainable with the DQL only.

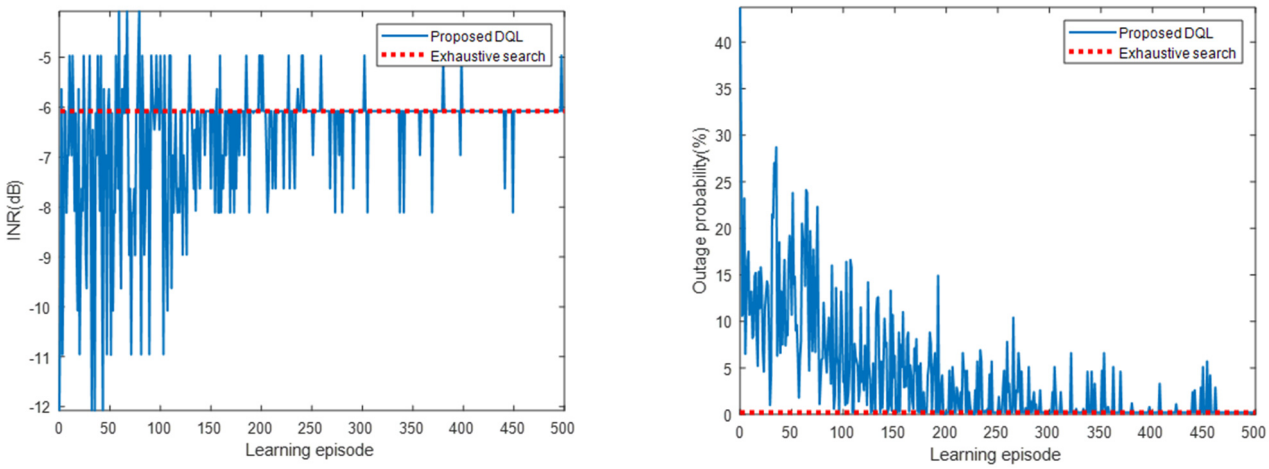


Figure 13. (**a**) INR and (**b**) p_{out} for each learning episode for interfered receiver ②.

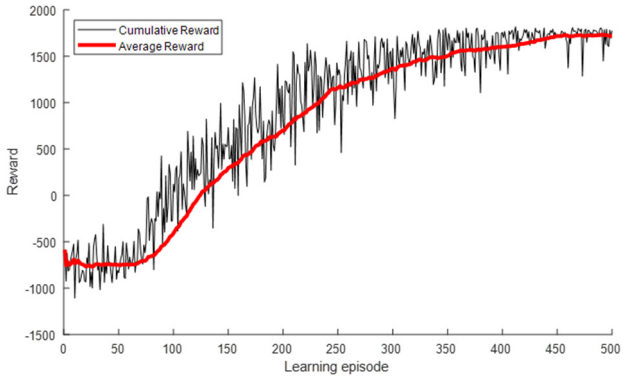


Figure 14. Reward for each learning episode for interfered receiver ②.

5.2.3. Simulation Results for Interfered Receiver ③

Figure 15 shows the SINR map based on $\boldsymbol{P}_c$ obtained using the proposed DQL-based power control algorithm for interfered receiver ③. The interfered receiver was located in the azimuth main lobe direction of *Cell*_3. It was closer to the HAPS than the receiver considered in Section 5.2.1 and was more severely affected by *Cell*_3; INR_{th} was not satisfied even for the minimum power of *Cell*_3. Thus, the optimal power control adjusted the power of *Cell*_2 and *Cell*_4, which caused the second-most interference. Table 6 presents a comparison of the $\boldsymbol{P}^*$ values obtained using an exhaustive search and $\boldsymbol{P}_c$ and a comparison of the DQL and DDQL results. Although the p_{out} of $\boldsymbol{P}_c$ was 0.6% higher than that of $\boldsymbol{P}^*$, it corresponded to the third-smallest value among the 78,125 values generated by the exhaustive search algorithm. In summary, the proposed power control algorithm achieved outstanding performance close to the optimal value.

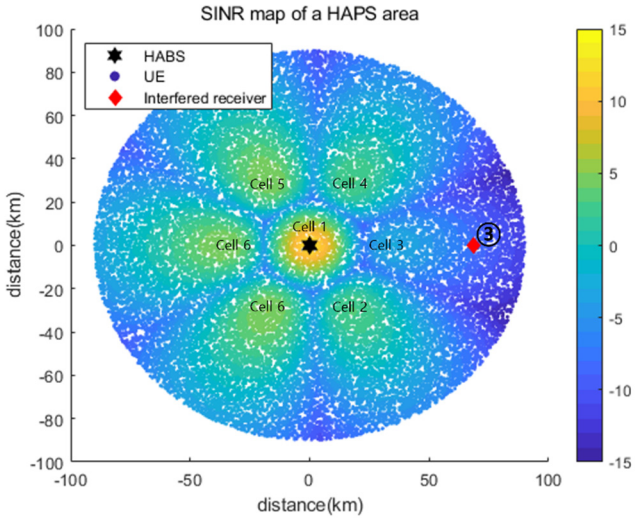


Figure 15. SINR map based on $\boldsymbol{P}_c$ obtained using the proposed DQL-based power control algorithm for interfered receiver ③.

Table 6. Performance comparison for interfered receiver ③.

	$P_{Cell\,1}$(dBm)	$P_{Cell\,2}$(dBm)	$P_{Cell\,3}$(dBm)	$P_{Cell\,4}$ (dBm)	$P_{Cell\,5}$ (dBm)	$P_{Cell\,6}$ (dBm)	$P_{Cell\,7}$ (dBm)	*INR* (dB)	P_{out} (%)
Optimal	37	34	26	32	34	34	34	−6.02	5.1
DQL	37	32	26	32	34	34	34	−6.06	5.7
DDQL	37	32	26	32	34	34	34	−6.06	5.7

As shown in Figure 16, the INR and p_{out} converged to the optimal values of the exhaustive search algorithm, with slight gaps. Similar to the results presented in Section 5.2.1, as the learning progressed, the INR converged from -6.19 to -6.06 dB, and the p_{out} converged from 43.7% to 5.7%. Figure 17 shows the cumulative and average rewards for each learning episode. The reward exhibited no noticeable improvement until approximately 130 episodes, after which it rapidly increased and then gradually converged at approximately 350 episodes. This is because to satisfy the INR_{th}, more agents had to take action, and the actions had to be more diverse. Nonetheless, the proposed DQL training process allowed the agent to learn the power control algorithm quickly and stably. We compared the learning results of the DQL and DDQL. Even when the DDQL was used, the results were the same as in Table 6 and Figures 16 and 17, verifying that the desired learning is attainable with the DQL only.

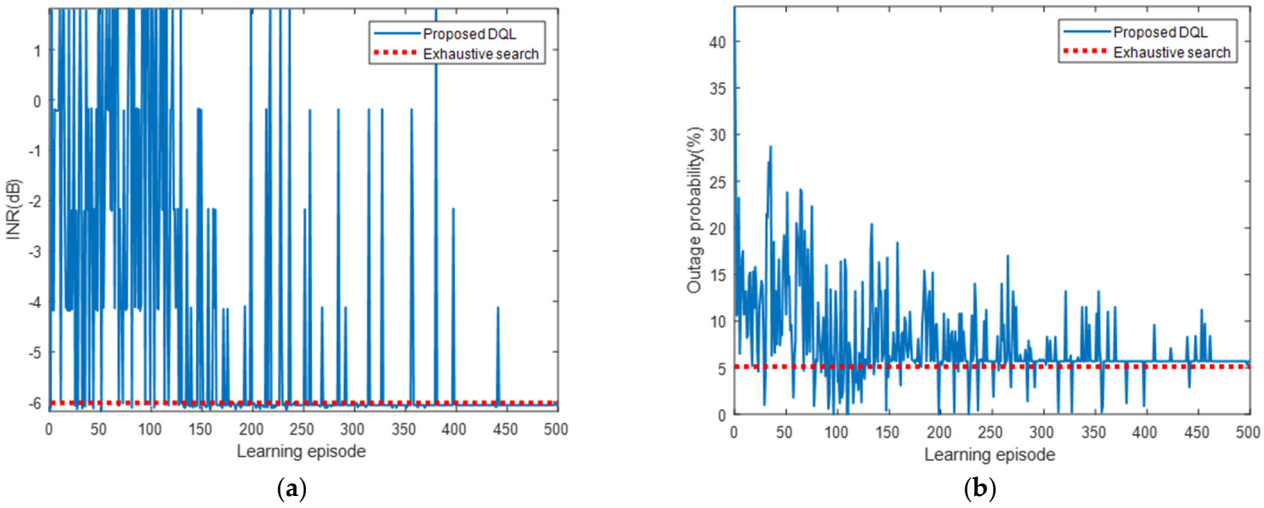


Figure 16. (**a**) INR and (**b**) p_{out} for each learning episode for interfered receiver ③.

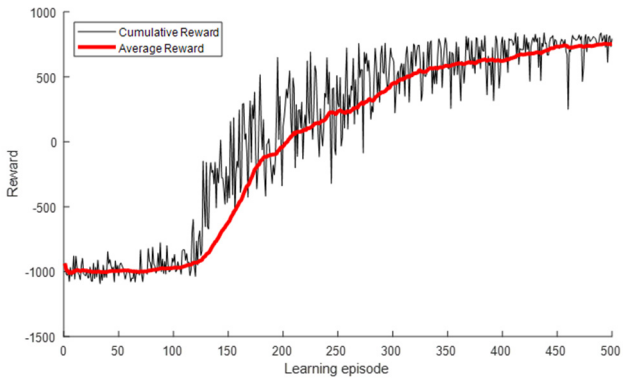


Figure 17. Reward for each learning episode for interfered receiver ③.

6. Conclusions

This paper proposed a DQL-based transmission power control algorithm for multicell HAPS communication that involved spectrum sharing with existing services. The proposed algorithm aimed to find a solution to the power control optimization problem for minimizing the outage probability of the HAPS downlink under the interference constraint to protect existing systems. We compared the solution with the optimal solution acquired using the exhaustive search algorithm. The simulation results confirmed that the proposed algorithm was comparable to the optimal exhaustive search.

Future work will include various power levels and expanding to multiple-HAPS communication in spectrum sharing with multiple interference systems. Since the increase

in the power level could reveal a value-based algorithm's limit, it is preferred to apply the policy-based algorithm. Given that multiple-HAPS communication could lead to the non-stationarity problem of multiagent reinforcement learning, its solution would be worth studying.

Author Contributions: Conceptualization and methodology, S.J. and H.-S.J.; software, S.J. and W.Y.; validation, formal analysis, and investigation, S.J. and W.Y. and H.-S.J.; resources and data curation, H.K.C. and E.N.; writing—original draft preparation, S.J. and W.Y.; writing—review and editing, S.J., H.-S.J. and J.P.; visualization, W.Y. and H.K.C.; supervision, J.P.; project administration, H.-S.J. and J.P.; funding acquisition, J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Agency for Defense Development (ADD).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Arum, S.C.; Grace, D.; Mitchell, P.D. A review of wireless communication using high-altitude platforms for extended coverage and capacity. *Comput. Commun.* **2020**, *157*, 232–256. [CrossRef]
2. Alam, M.S.; Kurt, G.K.; Yanikomeroglu, H.; Zhu, P.; Đào, N.D. High altitude platform station based super macro base station constellations. *IEEE Commun. Mag.* **2021**, *59*, 103–109. [CrossRef]
3. Kurt, G.K.; Khoshkholgh, M.G.; Alfattani, S.; Ibrahim, A.; Darwish, T.S.; Alam, M.S.; Yanikomeroglu, H.; Yongacoglu, A. A vision and framework for the high altitude platform station (HAPS) networks of the future. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 729–779. [CrossRef]
4. Hsieh, F.; Jardel, F.; Visotsky, E.; Vook, F.; Ghosh, A.; Picha, B. UAV-based Multi-cell HAPS Communication: System Design and Performance Evaluation. In Proceedings of the GLOBECOM 2020-2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020.
5. Xing, Y.; Hsieh, F.; Ghosh, A.; Rappaport, T.S. High Altitude Platform Stations (HAPS): Architecture and System Performance. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Online, 7–11 April 2021.
6. International Telecommunications Union (ITU). *World Radio Communication Conference 2019 (WRC-19) Final Acts*; International Telecommunications Union: Geneva, Switzerland, 2020; p. 366.
7. Ibrahim, A.; Alfa, A.S. Using Lagrangian relaxation for radio resource allocation in high altitude platforms. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 5823–5835. [CrossRef]
8. Guan, M.; Wu, Z.; Cui, Y.; Cao, X.; Wang, L.; Ye, J.; Peng, B. An intelligent wireless channel allocation in HAPS 5G communication system based on reinforcement learning. *EURASIP J. Wirel. Commun. Netw.* **2019**, *2019*, 1–9. [CrossRef]
9. Guan, M.; Wang, L.; Chen, L. Channel allocation for hot spot areas in HAPS communication based on the prediction of mobile user characteristics. *Intell. Autom. Soft Comput.* **2016**, *22*, 613–620. [CrossRef]
10. Oodo, M.; Miura, R.; Hori, T.; Morisaki, T.; Kashiki, K.; Suzuki, M. Sharing and compatibility study between fixed service using high altitude platform stations (HAPS) and other services in the 31/28 GHz bands. *Wirel. Pers. Commun.* **2002**, *23*, 3–14. [CrossRef]
11. Mokayef, M.; Rahman, T.A.; Ngah, R.; Ahmed, M.Y. Spectrum sharing model for coexistence between high altitude platform system and fixed services at 5.8 GHz. *Int. J. Multimed. Ubiquitous Eng.* **2013**, *8*, 265–275. [CrossRef]
12. Lee, W.; Jeon, Y.; Kim, T.; Kim, Y.I. Deep Reinforcement Learning for UAV Trajectory Design Considering Mobile Ground Users. *Sensors* **2021**, *21*, 8239. [CrossRef] [PubMed]
13. Koushik, A.M.; Hu, F.; Kumar, S. Deep Q-Learning-Based Node Positioning for Throughput-Optimal Communications in Dynamic UAV Swarm Network. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *5*, 554–566. [CrossRef]
14. Raja, G.; Anbalagan, S.; Narayanan, V.S.; Jayaram, S.; Ganapathisubramaniyan, A. Inter-UAV collision avoidance using Deep-Q-learning in flocking environment. In Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, New York, NY, USA, 10–12 October 2019; pp. 1089–1095.
15. Fotouhi, A.; Ding, M.; Hassan, M. Deep Q-Learning for Two-Hop Communications of Drone Base Stations. *Sensors* **2021**, *21*, 1960. [CrossRef] [PubMed]
16. Anicho, O.; Charlesworth, P.B.; Baicher, G.S.; Nagar, A.K. Reinforcement learning versus swarm intelligence for autonomous multi-HAPS coordination. *SN Appl. Sci.* **2021**, *3*, 1–11. [CrossRef]
17. Hasselt, H.V.; Guez, A.; Silver, D. Deep reinforcement learning with double q-learning. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.

18. International Telecommunications Union Radiocommunication Sector (ITU-R). *Working Document towards a Preliminary Draft New Report ITU-R M.[HIBS-CHARACTERISTICS]/Working Document Related to WRC-23 Agenda Item 1.4*; R19-WP5D Contribution 716 (Chapter 4-Annex 4.19); International Telecommunications Union: Geneva, Switzerland, 2021.
19. International Telecommunications Union Radiocommunication Sector (ITU-R). *Modelling and Simulation of IMT Networks and Systems for Use in Sharing and Compatibility Studies*; Recommendation ITU-R M.2101-0; International Telecommunications Union: Geneva, Switzerland, 2017.
20. International Telecommunications Union Radiocommunication Sector (ITU-R). *Reference Radiation Patterns of Omnidirectional, Sectoral and Other Antennas for the Fixed and Mobile Service for Use in Sharing Studies in the Frequency Range from 400 MHz to about 70 GHz*; Recommendation ITU-R F.1336-5; International Telecommunications Union: Geneva, Switzerland, 2019.
21. International Telecommunications Union Radiocommunication Sector (ITU-R). *Propagation Data Required for the Evaluation of Interference between Stations in Space and Those on the Surface of the Earth*; Recommendation ITU-R P.619-5; International Telecommunications Union: Geneva, Switzerland, 2021.
22. International Telecommunications Union Radiocommunication Sector (ITU-R). *Working Document towards Sharing and Compatibility Studies of HIBS under WRC-23 Agenda Item 1.4—Sharing and Compatibility Studies of High-Altitude Platform Stations as IMT Base Stations (HIBS) on WRC-23 Agenda Item 1.4*; R19-WP5D Contribution 716 (Chapter 4—Annex 4.20); International Telecommunications Union: Geneva, Switzerland, 2021.
23. Mnih, V.; Kavucuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
24. International Telecommunications Union Radiocommunication Sector (ITU-R). *Characteristics of Terrestrial Component of IMT for Sharing and Compatibility Studies in Preparation for WRC-23*; R19-WP5D Temporary Document 422 (Revision 2); International Telecommunications Union: Geneva, Switzerland, 2021.