Received 12 August 2023; revised 15 September 2023; accepted 2 October 2023. Date of publication 13 October 2023; date of current version 2 November 2023.

Digital Object Identifier 10.1109/OJCOMS.2023.3323547

Optimal Trajectory and Positioning of UAVs for Small Cell HetNets: Geometrical Analysis and Reinforcement Learning Approach

MOHAMMAD TAGHI DABIRI, MAZEN HASNA (Senior Member, IEEE), NIZAR ZORBA (Senior Member, IEEE), AND TAMER KHATTAB (Senior Member, IEEE)

Department of Electrical Engineering, Qatar University, Doha, Qatar

CORRESPONDING AUTHOR: M. T. DABIRI (e-mail: m.dabiri@qu.edu.qa)

This work was supported in part by the Qatar National Research Fund (QNRF) under Grant NPRP13S-0130-200200, and in part by the Qatar National Library.
This paper was presented in part at IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 2023, pp. 1–5 [1]
[DOI: 10.1109/VTC2023-Spring57618.2023.10199600].

ABSTRACT In this paper, a dynamic unmanned aerial vehicle (UAV)-based heterogeneous network (HetNet) equipped with directional terahertz (THz) antennas is studied to solve the problem of transferring massive traffic of distributed small cells to the core network. To this end, we first characterize a detailed three-dimensional (3D) modeling of the dynamic UAV-assisted HetNet, by taking into account the random positions of small cell base stations (SBSs), spatial angles between THz links, real antenna pattern, and UAV's vibrations in the 3D space. We then formulate the problem for UAV trajectory to minimize the maximum outage probability (OP) of directional THz links. Then, using geometrical analysis and deep reinforcement learning (RL) method, we propose several algorithms to find the optimal trajectory and select an optimal pattern during the trajectory. For a network with slow time changes, we also propose a deep RL framework to solve the joint optimal UAV positioning and antenna pattern control. The simulation results confirm that the UAV trajectory or antenna pattern control is not enough to achieve acceptable performance, and the UAV should control its antenna patterns during the trajectory to manage the interference.

INDEX TERMS Antenna pattern, deep reinforcement learning, positioning, trajectory, THz, UAV.

I. INTRODUCTION

Wireless backhaul/fronthaul links are proposed as an alternative for massive deployment of small cells because they are more flexible, easy to deploy, and cost-effective as compared to the traditional optical fiber links. Microwave backhaul/fronthaul links can cover a wide area but suffer from low data rates. High frequency millimeter wave (mmWave) and terahertz (THz) links meet the capacity requirements of next generation communication networks. However, mmWave/THz links suffer from susceptibility to weather conditions and require a line-of-sight (LoS) connection, which is the main hurdle in urban regions. A scalable idea was presented in [2] that utilizes unmanned aerial vehicles (UAVs) as a wireless fronthaul hub point between small cells and the core network. These UAV-hubs acting as networked flying platforms (NFPs) provide a possibility of wireless LoS fronthaul link and thus, enable the implementation of mmWave/THz in commercial systems.

However, in a dynamic network, the design of a UAV-based network with THz links is complicated because capacity and volume of demand for Internet access are dynamically changing during the day. The UAVs should adjust their positions in the three-dimensional (3D) space in relation to the distributed dense small cell base stations (SBSs) in such a way that, while providing a LoS connection, linklengths should be minimized to reduce channel loss and the spatial angle between links should be maximized to reduce interference between links. To this end, trajectory and positioning of UAV for providing high-quality THz fronthaul links for distributed dense small cell network is one of the main challenges for the implementation of THz backhaul/fronthaul in commercial systems; and this constitutes the main subject of this study. We tackle such challenge by combining geometrical analysis and reinforcement learning (RL) methods.

A. LITERATURE REVIEW

Although numerous great works have been reported about UAV trajectory in the literature [3], [4], [5], [6], [7], [8], most of these works are related to traditional RF frequency with omnidirectional antennas and cannot be directly used for directional THz antennas. Unlike omnidirectional antennas, directional THz antennas are sensitive to UAV's vibrations, alignment error and spatial angle between links. To solve the trajectory and positioning optimization problem for a 3D active topology, stochastic geometry can be applied. By studying survey [9] and the references there in, the main issue with this approach is that most of these works are based on very simplified assumptions that are not suitable for UAV-based 3D networks. For example, in most of the analysis, the focus is on the analysis of the distance between the user to the base station, which is suitable for common omnidirectional antennas in the access link. According to [10], [11], UAV-based THz/mmWave networks are function of distance, elevation and azimuth angles, and spatial angle between directional links. Most important, all these parameters are dependent on each other and it is not accurate to analyze the effect of each of these parameters separately for optimal network design. Therefore, the accurate analysis of such 3D systems using stochastic geometry analysis would be very complicated, if not impossible.

The complexity of stochastic geometry analysis for modern wireless networks are continuously increasing, and tools from artificial intelligence (AI) and machine learning (ML) are crucial [12]. UAV trajectory/positioning with the help of RL algorithms can provide a reliable service for distributed users, which is the subject of several recent works [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. In general, omnidirectional antennas have been used in all these works. For omnidirectional antennas, the problem is very simple because only the distance of the links is important. For THz directional antennas, the distribution of ground nodes, and the spatial angle between adjacent nodes affect the channel and interference model and thus, the trajectory optimization algorithms provided in these works cannot be directly used. Moreover, due to the small beamwidth of directional THz links, the small fluctuations of the UAV, even in the order of one degree, can affect the performance of the system, and therefore the THz beamwidth have to be designed with this constraint in mind, and avoiding small beamwidth values. Large beamwidth values, on the other hand, allow interference among THz links. Finding the suitable adjustment point has to be dynamic during the trajectory, where the UAV must simultaneously adjust its antenna pattern to control the interference between randomly distributed nodes which is the subject of this work.

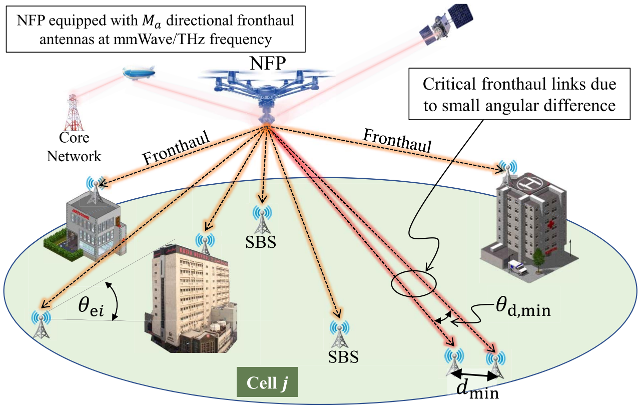

NFP equipped with M_a directional fronthaul antennas at mmWave/THz frequency
NFP
Critical fronthaul links due to small angular difference
Core Network
Fronthaul
Fronthaul
SBS
SBS
θ_{ei}
$\theta_{d,\min}$
Cell j
$d_{\min}$

FIGURE 1. An illustration of a UAV-assisted HetNet as an alternative solution for fronthaul links which uses directional THz antenna to transfer traffic from the distributed SBSs to the core network.

B. CONTRIBUTIONS

In this study, we consider a dynamic UAV-assisted HetNet as shown in Fig. 1 that is offered as a cost effective and easy to deploy solution in [2] to solve the problem related to transferring traffic of the distributed SBSs to the core network. In order to increase the network capacity, we use THz directional antennas to reuse the frequency in the 3D space. The most important challenge of the considered network is the interference between the links. Therefore, in order to manage the interference, it is necessary for the UAV to perform its trajectory in such a way that, while placing all the SBSs in its field-of-view (FoV), it maximizes the angle between the links in order to achieve the least interference between the links. Also, during the trajectory, the antenna patterns should be managed in such a way that they are resistant to the UAV's fluctuations while reducing interference. Designing the considered dynamic network with the mentioned challenges is the main goal of this work. To this end, by combining geometrical analysis and deep RL, we propose several novel algorithms for UAV trajectory and optimal antenna pattern control. Our main contributions are summarized as follows:

- We characterize a detailed 3D modeling of the dynamic UAV-assisted HetNet, by taking into account the random positions of SBSs, spatial angles between THz links, real antenna patterns, and UAV's vibrations in the 3D space. Using this characterization, we formulate the problem for two scenarios including UAV trajectory for a dynamic networks and positioning for a network with slow time changes.
- For the simple case of constant antenna pattern along the trajectory, a deep RL framework is proposed to solve the trajectory problem of the dynamic network with the objective of minimizing the maximum outage probability (OP) of fronthaul links.
- To tackle the high complexity of joint optimal UAV trajectory and antenna pattern beamwidth control, a novel method is provided based on the combination of geometrical analysis and deep RL method.

- For a network with slow time changes, we also propose an algorithm to solve the joint optimal UAV positioning and antenna pattern control.
- Then, using the geometrical analysis of the random positions of SBSs, we provide a novel algorithm for UAV-positioning that converges to the global optimal position very fast.
- Next, we examine the performance of the system in the trajectory and positioning scenarios, through both algorithms evaluation and computer simulations.
- Finally, we modify our proposed algorithms for an environment with 3D obstacles and we show that by using the proposed algorithms, the UAV learns well during the trajectory to first place all the SBSs in the LoS state. Then, with the optimal antenna pattern selection, the UAV finds the optimal trajectory to reduce the interference between the SBSs that are close to each other.

II. THE SYSTEM MODEL

As shown in Fig. 1, we consider a UAV-assisted HetNet where its UAV-based THz links are used to transfer the SBSs traffic. We assume each UAV is equipped with M_u directional antennas denoted by A_i where $i \in \{1, \ldots, M_u\}$. The direction of each A_i is set towards an SBS assigned to it. SBSs are randomly distributed in a two-dimensional ground space and they can change randomly over time. Let $M_s \in \{M_{\text{s,min}}, \ldots, M_{\text{s,max}}\}$ denote the number of SBSs which is a uniform discrete random variable (RV) where $M_{\text{s,max}} = M_u$. The position of each SBS denoted by S_i is characterized as $[x_i, y_i, 0]$ in a Cartesian coordinate system where $[x, y, z] \in \mathbb{R}^{1\times 3}$. The SBS adjusts its directional antenna (denoted by A_{s_i}) towards the UAV. We assume that the ground SBS has higher stability than the UAV and has a negligible vibration error compared to the UAV. Also, F_i stands for the fronthaul link between A_i and A_{s_i}.

We consider a dynamic general network in which the number and position of SBSs change randomly with time $T \in [T_{\min}, T_{\max}]$ which is a uniform RV. The changes in the network after time T are such that m_d of the SBSs are randomly disconnected and the other m_c SBSs are connected to the UAV with new random positions on the $x - y$ plane.

With any change in network topology, the UAV must continuously modify its position. We assume that the UAV flies with the maximum speed constraint $v_{\max}$ in (m/s) and the maximum acceleration constraint $v'_{\max}$ in (m/s^2). Let $B_u(t) = [x_u(t), y_u(t), z_u(t)]$ represent the instantaneous position of UAV at time t. For a short time Δt, the position of UAV is updated as

$$\begin{cases} x_u(t+\Delta t) = x_u(t) + v_x(t)\Delta t + \frac{1}{2}v'_x(t)\Delta t^2, \\ y_u(t+\Delta t) = y_u(t) + v_y(t)\Delta t + \frac{1}{2}v'_y(t)\Delta t^2, \\ z_u(t+\Delta t) = z_u(t) + v_z(t)\Delta t + \frac{1}{2}v'_z(t)\Delta t^2, \end{cases} \tag{1}$$

where $v(t) = [v_x(t)], v_y(t), v_z(t)]$ and $v'(t) = [v'_x(t)], v'_y(t), v'_z(t)]$ are instantaneous speed and acceleration of UAV, respectively. For notation simplicity, we will remove the notation t of $B_u(t) = [x_u(t), y_u(t), z_u(t)]$ in the following, except where necessary.

A. THE 3D ANTENNA PATTERN

In order to reduce the effect of interference and also to reduce the negative effect of channel attenuation at high THz frequencies, the use of high gain antenna is essential, particularly for ultra high data rate fronthaul links. In addition, as shown in [11], employing a directional antenna pattern allows us to reuse frequency bands thus improving the spectral efficiency of the considered system.

We consider a uniform square array antennas for both UAV and SBSs. Let $N_{ui} \times N_{ui}$ represent antenna elements of A_i for $i \in \{1, \ldots, M_u\}$ with the same spacing d_a between elements. Similarly, $N_{si} \times N_{si}$ are antenna elements of S_i for $i \in \{1, \ldots, M_s\}$. The array radiation gain is mainly formulated in the direction of θ and ϕ, where θ and ϕ are clearly defined in [26, Fig. 6.28]. By taking into account the effect of all elements, the array radiation gain will be:

$$G_{qi}\big(N_{qi}, \theta, \phi\big) = G_0\big(N_{qi}\big) G_{ai}\big(N_{qi}, \theta, \phi\big), \tag{2}$$

where G_{ai} is an array factor and G_0 is defined in (4). Also, the subscript $q = s$ determines the antenna of S_i and the subscript $q = u$ determines the antenna of A_i. If the amplitude excitation of the entire array is uniform, then the array factor $G_{ai}(N_{qi}, \theta, \phi)$ for a square array of $N_{qi} \times N_{qi}$ elements can be obtained as [26, eqs. (6-89) and (6-91)]:

$$G_{ai}\big(N_{qi}, \theta, \phi\big) = \left(\frac{\sin\left(\frac{N_{qi}(kd_a \sin(\theta)\cos(\phi) + \mathbb{V}_x)}{2}\right)}{N_{qi}\sin\left(\frac{kd_a \sin(\theta)\cos(\phi) + \mathbb{V}_x}{2}\right)} \times \frac{\sin\left(\frac{N_{qi}(kd_a \sin(\theta)\sin(\phi) + \mathbb{V}_y)}{2}\right)}{N_{qi}\sin\left(\frac{kd_a \sin(\theta)\sin(\phi) + \mathbb{V}_y}{2}\right)} \right)^2, \tag{3}$$

where $d_a = \frac{\lambda}{2}$ and $\mathbb{V}_w$ are the spacing and progressive phase shift between the elements, respectively. $k = \frac{2\pi}{\lambda}$ is the wave number, $\lambda = \frac{c}{f_c}$ is the wavelength, f_c is the carrier frequency, and c is the speed of light. Also, in order to guarantee that the total radiated power of antennas with different N_{qi} are the same, the coefficient G_0 is defined as

$$G_0\big(N_{qi}\big) = \frac{4\pi}{\int_0^{\pi}\int_0^{2\pi} G_{ai}\big(N_{qi}, \theta, \phi\big)\sin(\theta)d\theta d\phi}. \tag{4}$$

Based on (3), the maximum value of the antenna gain is equal to $G_0(N_{qi})$, which is obtained when $\theta = 0$.

B. MODELING OF UAV VIBRATIONS

In practical situations, an error in mechanical control system, mechanical noise, position estimation errors, air pressure, and wind speed can affect the UAV's angular and position stability [27]. This, in turn, leads to antenna misalignment or pointing errors. Therefore, as N_{qi} increases, the system becomes more sensitive to the UAV's vibrations. On the one hand, we must increase N_{qi} to reduce interference between fronthaul links. However, we must be careful not to choose a very large value for N_{qi} such that even with small fluctuations in antenna direction, the probability of missing the main lobe and being on the side-lobes increases. Let $\Theta = [\Theta_x, \Theta_y]$ denote the UAV's orientation fluctuations. Based on the central limit theorem, the UAV's orientation fluctuations are considered to be Gaussian distributed [28], [29]. Therefore, we have $\Theta_x \sim \mathcal{N}(0, \sigma_\theta^2)$, and $\Theta_y \sim \mathcal{N}(0, \sigma_\theta^2)$. The received power by A_{s_i} is modeled as follows:

$$\begin{aligned} P_{r_i} &= P_{t_i}|h_{L_i}|^2 G_0(N_{si}) G_{uj}\big(N_{uj}, \Theta, \Phi\big) \\ &\quad + \underbrace{|h_{L_i}|^2 \sum_{j=1, j\neq i}^{M_s} P_{t_j} G_0(N_{si}) G_{uj}\big(N_{uj}, \theta_{ij}, \phi_{ij}\big)}_{\text{Interference}} + n_0, \end{aligned} \tag{5}$$

where n_0 is the receiver noise, P_{t_i} is the transmit power of A_i, L_i is the link length of F_i, $h_{L_i} = h_{Lf}(L_i) h_{Lm}(L_i)$ is the channel path loss, $h_{Lf}(L_i) = (\frac{\lambda}{4\pi L_i})^2$ is the free-space path loss, $h_{Lm}(L_i) = e^{-\frac{\mathcal{K}(f)}{2} L_i}$ represents the molecular absorption loss, and $\mathcal{K}(f)$ is the frequency dependent absorption coefficient. Moreover, in (5), $\theta_{ij} = [\theta_{x_{ij}} + \Theta_x, \theta_{y_{ij}} + \Theta_y]$, where $\theta'_{ij} = [\theta_{x_{ij}}, \theta_{y_{ij}}]$ is the spatial angle between F_i and F_j links which is obtained as

$$\begin{cases} \theta_{x_{ij}} = \cos^{-1}\left(\dfrac{(x_u - x_i)^2 + (x_u - x_j)^2 + 2z_u^2 - d_{x_{ij}}^2}{2\sqrt{\left[(x_u - x_i)^2 + z_u^2\right]\left[(x_u - x_j)^2 + z_u^2\right]}}\right), \\ \theta_{y_{ij}} = \cos^{-1}\left(\dfrac{(y_u - y_i)^2 + (y_u - y_j)^2 + 2z_u^2 - d_{y_{ij}}^2}{2\sqrt{\left[(y_u - y_i)^2 + z_u^2\right]\left[(y_u - y_j)^2 + z_u^2\right]}}\right), \end{cases} \tag{6}$$

where $d_{x_{ij}} = |x_i - x_j|$ and $d_{y_{ij}} = |y_i - y_j|$. Also, the parameter ϕ_{ij} is the roll angle of pattern A_j with respect to A_{s_i}.

C. PROBABILITY OF LoS

In addition to the high propagation loss, THz communication systems are very sensitive to blockages [30]. Therefore, the probability of LoS is an important factor and can be described as a function of the elevation angle and environment as follows [31], [32]:

$$P_{\text{LoS}}(\theta_{ei}) = \frac{1}{1 + \alpha \exp\Big(-b\Big(\frac{180}{\pi}\theta_{\text{ei}} - \alpha\Big)\Big)} \tag{7}$$

where α and b are constants whose values depend on the propagation environment, e.g., rural, urban, or dense urban, and θ_{ei} is the elevation angle of S_i compared to the instantaneous position of UAV and can be formulated as

$$\theta_{ei} = \tan^{-1}\left(\frac{z_u}{\sqrt{(x_i - x_u)^2 + (y_i - y_u)^2}}\right). \tag{8}$$

Finally, the SINR is modeled as

$$\gamma_i = \frac{P_{t_i}\alpha_{L_i}|h_{L_i}|^2 G_0(N_{si}) G_{uj}\big(N_{uj}, \Theta, \Phi\big)}{\sum_{j=1, j\neq i}^{M_s} P_{t_j}\alpha_{L_i}|h_{L_i}|^2 G_0(N_{si}) G_{uj}\big(N_{uj}, \theta_{ij}, \phi_{ij}\big) + \sigma_N^2}, \tag{9}$$

where σ_N^2 is the thermal noise power, and coefficient α_{L_i} determines S_i is in the LoS or non-line-of-sight (NLoS) of the UAV.

III. PROBLEM FORMULATION

As mentioned, the distribution of active fronthaul links connected to the UAV changes with time T, which is a random parameter. If T is in the order of a few seconds, our problem is of the trajectory type, because the UAV must continuously correct its position. If T changes in the order of several tens of seconds to minutes, the UAV has enough time to reach the optimal position, and therefore our problem is of the positioning type. In practice, the topology of active/inactive fronthaul links changes less than access links, and this assumption is also practical for many scenarios.

Let us represent a set of UAV movements in the time period $\mathcal{T} = [t_0, t_0 + J_a \Delta t]$ as

$$\mathcal{B}_u(\mathcal{T}) = \{B_u(t_0), B_u(t_0 + \Delta t), \ldots, B_u(t_0 + J_a \Delta t)\},$$

where $B_u(t + j\Delta t) = [x_u(t + j\Delta t), y_u(t + j\Delta t), z_u(t + j\Delta t)]$ for $j \in \{0, 1, \ldots, J_a\}$, and J_a is the number of UAV's actions. $\mathcal{N}'_u(\mathcal{T})$ is defined as a set of antenna patterns during the time period $\mathcal{T}$ as

$$\mathcal{N}'_u(\mathcal{T}) = \{\mathcal{N}_u(t_0), \mathcal{N}_u(t_0 + \Delta t), \ldots, \mathcal{N}_u(t_0 + J_a \Delta t)\}, \tag{10}$$

where $\mathcal{N}_u(t_0 + j\Delta t)$ is the set of active antenna elements at time $t = t_0 + j\Delta t$ for M_s different directional A_i antennas which is formulated as

$$\mathcal{N}_u(t_0 + j\Delta t) = \big\{N_{u1}, N_{u2}, \ldots, N_{uM_s}\big\}, \tag{11}$$

for $j \in \{0, 1, \ldots, J_a\}$. Also, $\mathcal{P}'_t(\mathcal{T})$ is defined as a set of transmitted power during the time period $\mathcal{T}$ as

$$\mathcal{P}'_t(\mathcal{T}) = \{\mathcal{P}_t(t_0), \mathcal{P}_t(t_0 + \Delta t), \ldots, \mathcal{P}_t(t_0 + J_a \Delta t)\}, \tag{12}$$

where $\mathcal{P}_t(t_0 + j\Delta t)$ is the set of antenna patterns at time $t = t_0 + j\Delta t$ for M_s different A_i antennas which is formulated as

$$\mathcal{P}_t(t_0 + j\Delta t) = \big\{P_{t1}, P_{t2}, \ldots, P_{tM_s}\big\}, \tag{13}$$

for $j \in \{0, 1, \ldots, J_a\}$.

The trajectory time is denoted as $\mathbb{T}_{ep}$ and is defined as:

$$\mathbb{T}_{ep} = \sum_{j=1}^{J_a} \frac{\Delta B_u(t_0 + j\Delta t)}{v(t_0 + j\Delta t)}, \tag{14}$$

where

$$\begin{aligned} &\Delta B_u(t_0 + j\Delta t) \\ &\quad = \sqrt{\Delta^2 x_u(t_0 + j\Delta t) + \Delta^2 y_u(t_0 + j\Delta t) + \Delta^2 z_u(t_0 + j\Delta t)}, \end{aligned}$$

and $\Delta x_u(t_0 + j\Delta t) = x_u(t + (j+1)\Delta t) - x_u(t + j\Delta t)$, $\Delta y_u(t_0 + j\Delta t) = y_u(t + (j+1)\Delta t) - y_u(t + j\Delta t)$, and $\Delta z_u(t_0 + j\Delta t) = z_u(t + (j+1)\Delta t) - z_u(t + j\Delta t)$. The trajectory time is more important because we have a dynamic network where on average, the topology of the network changes every $\bar{T} = \frac{T_{\max} + T_{\min}}{2}$ second. Regarding the trajectory problem, the

UAV seeks to find the optimal trajectory that will satisfy the requested QoS of all fronthaul links in a minimum time $\mathbb{T}_{ep}$ under the constraint:

$$\mathbb{T}_{ep} \leq \bar{T}. \tag{15}$$

Considering the importance of the OP in wireless communication, especially for wireless fronthaul links, in this work, we define the quality of service based on the OP [33]. To achieve fair performance among all fronthaul links, we minimize the maximum OP over all SBSs as

$$\min_{\mathcal{B}_u(\mathcal{T}),\mathcal{N}'_u(\mathcal{T}),\mathcal{P}'_t(\mathcal{T})} \max\left[P_{\text{out},1}, P_{\text{out},2}, \ldots, P_{\text{out},M_s}\right], \tag{16}$$

where $P_{\text{out},i}$ is the OP of the ith fronthaul link (i.e., the probability that γ_i falls below a threshold γ_{th}) and is obtained as [34]

$$\mathbb{P}_{\text{out},i} = \text{Prob}\left[\gamma_i < \gamma_{\text{th}}\right]. \tag{17}$$

Finally, based on (15) and (16), our optimization problem for trajectory is formulated as follows:

$$\begin{aligned}
&\min_{\mathcal{B}_u(\mathcal{T}),\mathcal{N}'_u(\mathcal{T}),\mathcal{P}'_t(\mathcal{T})} \max\left[P_{\text{out},1}, \ldots, P_{\text{out},M_s}\right] && \text{(18a)}\\
&\min_{\mathcal{B}_u(\mathcal{T}),\mathcal{N}'_u(\mathcal{T}),\mathcal{P}'_t(\mathcal{T})} \frac{1}{J_a}\sum_{j=0}^{J_a} \mathcal{P}_t(t_0 + j\Delta t) && \text{(18b)}\\
&\text{s.t.}\ \mathbb{T}_{ep} \leq \bar{T}, && \text{(18c)}\\
&P_{\text{out},i} < P_{\text{out,th}}, \quad i \in \{1, \ldots, M_s\}, && \text{(18d)}\\
&N_{\min} \leq N_{ui} \leq N_{\max}, \quad i \in \{1, \ldots, M_s\}, && \text{(18e)}\\
&P_{ti} \leq P_{t,\max}, \quad i \in \{1, \ldots, M_s\}, && \text{(18f)}\\
&h_{\min} \leq z_u(t) \leq h_{\max}, && \text{(18g)}\\
&v'(t) \leq v'_{\max}, && \text{(18h)}\\
&v(t) \leq v_{\max}. && \text{(18i)}
\end{aligned}$$

For a network with slow time changes, the trajectory time $\mathbb{T}_{ep}$ is not very important and our problem becomes positioning. The optimization problem for positioning is simplified as:

$$\begin{aligned}
&\min_{B_u,\mathcal{N}_u,\mathcal{P}_t} \max\left[P_{\text{out},1}, \ldots, P_{\text{out},M_s}\right] && \text{(19a)}\\
&\min_{B_u,\mathcal{N}_u,\mathcal{P}_t} \frac{1}{J_a}\sum_{j=0}^{J_a} \mathcal{P}_t(t_0 + j\Delta t) && \text{(19b)}\\
&\text{s.t.}\ P_{\text{out},i} < P_{\text{out,th}}, \quad i \in \{1, \ldots, M_s\}, && \text{(19c)}\\
&N_{\min} \leq N_{ui} \leq N_{\max}, \quad i \in \{1, \ldots, M_s\}, && \text{(19d)}\\
&P_{ti} \leq P_{t,\max}, \quad i \in \{1, \ldots, M_s\}, && \text{(19e)}\\
&h_{\min} \leq z_u(t) \leq h_{\max}, && \text{(19f)}\\
&v'(t) \leq v'_{\max}, && \text{(19g)}\\
&v(t) \leq v_{\max}. && \text{(19h)}
\end{aligned}$$

It should be noted that in the optimization problem (19), the optimal values of the parameters $\mathcal{B}_u(\mathcal{T})$, $\mathcal{N}'_u(\mathcal{T})$, and $\mathcal{P}'_t(\mathcal{T})$ along the trajectory are no longer important for us, and we are only looking for the optimal values of the parameters B_u, $\mathcal{N}_u$, and $\mathcal{P}_t$ at the end point of the trajectory.

IV. ANALYSIS AND ALGORITHMS

As we can see, the optimization problem in (18) for trajectory and its simplified version in (19) for positioning are NP-hard because they are nonconvex, nonlinear, and mixed discrete optimization problems [35]. Therefore, we are not able to solve the optimization problems in (18) and (19) by classical programming methods and hence, we move to use RL-based methods. In the following, we first focus on the trajectory optimization problem (18) and then, using the obtained results, we find an efficient solution for the positioning problem (19).

A. PRELIMINARIES ON PROPOSED DEEP RL-BASED METHODS

In order to select an appropriate RL method, we need to have an accurate knowledge about the state and action of our problem. We are targeting to find the optimal trajectory/position of the UAV and the antenna patterns along with the optimal power allocation in such a way that the interference between the fronthaul links is minimized. Therefore, our state space includes a continuous 3D space for the UAV's position and a continuous M_s-dimensional space for optimal power allocation along with a discrete M_s-dimensional space for the M_s antennas related to the M_s active fronthaul links. In this section, we will first solve the problem for the simple case by considering the constant antenna pattern and power, and then, we use the obtained results to solve the problem for the general case. In the simple case, the state space is a continuous 3D space for the position of the UAV, $s_t = B_u(t) = [x_u(t), y_u(t), z_u(t)]$ where $h_{\min} \leq z_u(t) \leq h_{\max}$, and the action is also a continuous 3D variable $a_t = [a_{xt}, a_{yt}, a_{zt}]$ where

$$s_{t+1} = \begin{cases} s_t + a_t, & \text{if} \quad h_{\min} < z_u(t) + a_{zt} < h_{\max}, \\ s_t, & \text{otherwise.} \end{cases} \tag{20}$$

Because the action space is continuous, gradient-based learning algorithms allow us to find the best parameters by just following the gradient. Thereby, for the considered UAV-based system, deep deterministic policy gradient (DDPG) algorithm or its variants are fit to find the optimal policy for the agent. While DDPG can achieve great performance, it is frequently unstable with respect to hyperparameters because there is a risk of overestimating Q-values in the critic (value) network [36]. Twin Delayed DDPG (TD3) is an efficient policy gradient algorithm that addresses this issue by introducing several critical tricks [36]. In this paper, TD3 is used to find an optimal policy for the continuous actions of UAV. Unlike the basic structure of an actor-critic network, TD3 consists of two critic deep neural networks (DNNs) $Q(s_t, a_t, \psi_i)$ for $i \in \{1, 2\}$, two target DNNs $Q(s_t, a_t, \psi'_i)$ related to the $Q(s_t, a_t, \psi_i)$, one actor DNN $\pi(s_t, \phi)$, and one target DNN $\pi(s_t, \phi'_i)$ related to the $\pi(s_t, \phi_i)$. Using square Bellman error minimization, TD3 concurrently learns two critic DNNs. The minimum of two similar critic $Q(s_t, a_t, \psi_i)$ is used to approximate the target Q-value. At every time training step, TD3 updates the parameters of each critic by minimizing the cost function for training the critic DNN which is defined as

$$\begin{cases} J_{\psi_1} = \underset{s_t,a_t,r_t,s_{t+1}}{\mathbb{E}_{\psi_1}} \left[(y_{\text{tar}} - Q(s_t, a_t, \psi_1))^2\right], \\ J_{\psi_2} = \underset{s_t,a_t,r_t,s_{t+1}}{\mathbb{E}_{\psi_2}} \left[(y_{\text{tar}} - Q(s_t, a_t, \psi_2))^2\right], \end{cases} \tag{21}$$

where

$$\begin{cases} y_{\text{tar}} = r_t(s_t, s_{t+1}, a_t) + \gamma \min_{i=1,2} Q\big(s_t, \tilde{a}_t, \psi_i'\big), \\ \tilde{a}_t = \text{clip}\big[\pi\left(s_{t+1}, \phi'\right) + \text{clip}[\epsilon, -\epsilon_{\min}, \epsilon_{\max}], a_{\text{Low}}, a_{\text{High}}\big], \end{cases} \tag{22}$$

and γ is the discount factor, $a_{\text{Low}} < a_t < a_{\text{High}}$ is the valid action range, and $\text{clip}[\epsilon, -\epsilon_{\min}, \epsilon_{\max}]$ is the clipped noise. Instead of running an expensive optimization subroutine each time to learn a deterministic policy $\pi(s, \phi)$, we can approximate $\max_a Q(s, a; \psi_1) \approx Q(s, \pi(s, \phi); \psi_1)$ [36]. Based on that, every d_1 steps, we update the parameters of actor ϕ by minimizing the following cost function:

$$J_\phi = \sum_s d_\pi(s) Q(s, \pi(s_t, \phi) + \epsilon; \psi_1), \tag{23}$$

where $a_t = \pi(s_t, \phi) + \epsilon$ is the final deterministic and continuous action, ϵ is added noise for exploration, and $d_\pi(s)$ is the state distribution. For our problem, since many episodes finish without achieving the goal state, we use a hindsight experience replay (HER) memory to enhance efficiency of the TD3 algorithm, wherein, we reset the final state of each episode instead of the goal state [37].

B. TD3-BASED TRAJECTORY WITH FIXED ANTENNA PATTERNS

To proceed with the TD3, we need to define state, action, and reward, described as follows.

1) REWARD

A TD3 agent is an actor-critic reinforcement learning agent that searches for an optimal policy that maximizes the expected cumulative long-term reward. The goal is to find the optimal UAV's trajectory in such a way that the OP in the worst conditions (fronthaul link with the worst outage probability) is minimized. Therefore, we define the reward as follows:

$$r_t = P_{\text{out}} = \max\big[P_{\text{out},1}, \ldots, P_{\text{out},M_s}\big]. \tag{24}$$

Usually, logarithmic scales are used to show the OP, because we can better see the performance improvement, especially for the lower OPs. Accordingly, in order to better recognize the difference between system performance at low OPs, we modify the reward as follows:

$$r_t = -\ln(P_{\text{out}}). \tag{25}$$

To calculate the OP, M_s different two-dimensional integrals need to be numerically solved which causes a long processing time to solve the problem. In order to decrease the run time of the optimization problem, closed-form expression for the OP is derived in Theorem 1.

Theorem 1: OP of the ith fronthaul link is derived as

$$\begin{aligned} \mathbb{P}_{\text{out},i} \simeq{}& Q\left(\frac{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2 - w_B^2(N_{ui}) \ln(\gamma_{\text{th}})}{2\sigma_\theta \sqrt{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2}}\right) P_{\text{LoS}}(\theta_{ei}) \\ &+ \left(\frac{\sigma_N^2 \gamma_{\text{th}}}{P_{t_i} |h_{L_i}|^2 G_0(N_{si}) G_0(N_{ui})}\right)^{\frac{w_B^2(N_{ui})}{2\sigma_\Theta^2}} \\ &\times U\left(1 - \frac{\sigma_N^2 \gamma_{\text{th}}}{P_{t_i} |h_{L_i}|^2 G_0(N_{si}) G_0(N_{ui})}\right) P_{\text{LoS}}(\theta_{ei}) \\ &+ (1 - P_{\text{LoS}}(\theta_{ei})), \end{aligned} \tag{26}$$

where j represents the fronthaul link with the lowest spatial angle compared to the fronthaul link i, and $U(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$. is the step function.

Proof: Please refer to Appendix A. ■

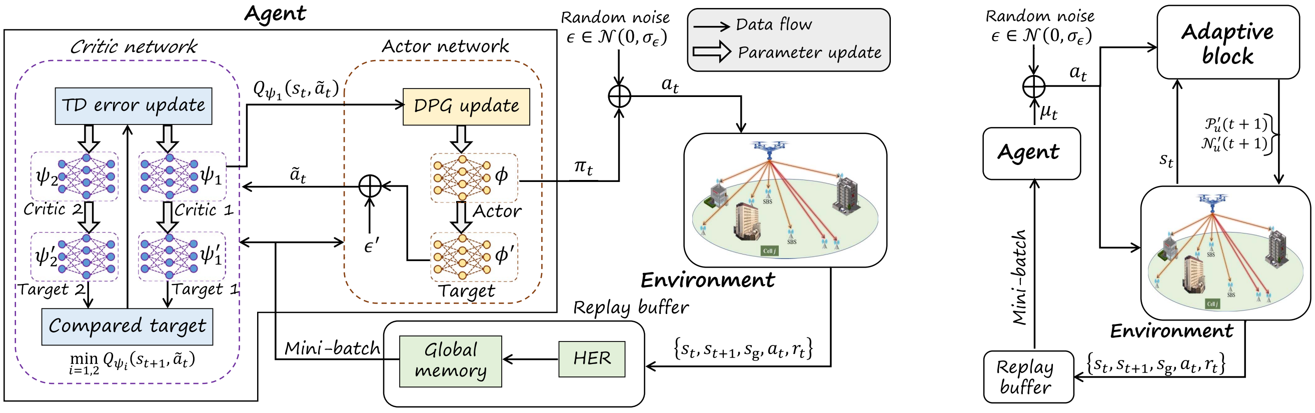
As can be seen from (26), the OP is computed from the sum of three terms. If the instantaneous spatial angle $\theta_{ij} = [\theta_{x_{ij}}, \theta_{y_{ij}}]$ is small, the OP of the F_i link is limited by interference of the F_j link. Otherwise, if the special angles between all the links are large enough and all the SBSs are in the FoV of the UAV, the OP is limited by the link with the longest link length, which is provided in second term of (26). By reducing the elevation angle θ_{ei}, which is a function of the instantaneous position of the UAV relative to the S_i, the OP of the F_i link will be limited by the third term of (26).

In Fig. 3, we examined the accuracy of (26) through simulations. In this figure, in addition to the total OP, we investigate the effects of the aforementioned three terms. The most important tunable parameter in the considered system model is the antenna pattern which can be adjusted with N_{ui}. The optimal selection of N_{ui} is a function of the θ_{ij} and the intensity of the UAV's instabilities. For this reason, the OP is plotted for a wide range of N_{ui}. As expected, for smaller N_{ui}, the beamwidth of the antenna pattern is enlarged and the system performance is limited by interference. By increasing N_{ui}, the beamwidth decreases and as a result the interference decreases, and the system is limited by noise. However, due to the UAV's vibrations, for larger N_{ui}, the beamwidth becomes very small, and as a result, the performance becomes more sensitive to the antenna alignment errors. Also, the simulation results confirm the accuracy of the analytical results.

2) STATE AND ACTION

According to Section II, the state is actually the instantaneous position of the UAV in 3D space denoted by $s(t) = [x_u(t), y_u(t), z_u(t)]$, which is a 3D continuous variable. Also, using (1), the action is a 3D continuous variable as

$$\begin{cases} a_x(t) = v_x(t)\Delta t + \frac{1}{2} v_x'(t) \Delta t^2, \\ a_y(t) = v_y(t)\Delta t + \frac{1}{2} v_y'(t) \Delta t^2, \\ a_z(t) = v_z(t)\Delta t + \frac{1}{2} v_z'(t) \Delta t^2. \end{cases} \tag{27}$$



(a) (b)

FIGURE 2. Structure of TD3-based algorithm for (a) trajectory planning, (b) trajectory planing with optimal pattern and power allocation.

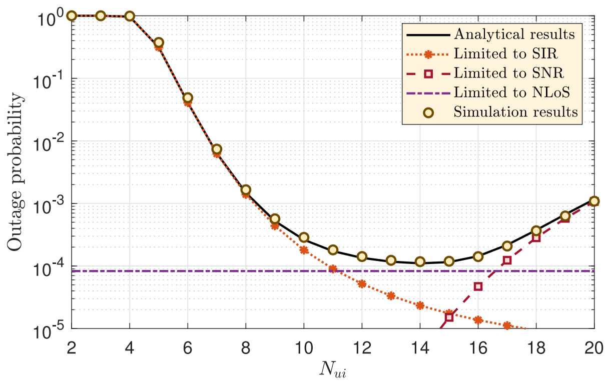


FIGURE 3. Comparing the accuracy of the analytical results for $P_{\text{out},i}$ provided in Theorem 1 with the simulation results for $\theta_{ij} = 18°$, $\sigma_\theta = 2°$ and different values of N_{ui}.

It can be easily shown that for any random positioning of SBSs, the optimal position of the UAV is in the following interval:

$$\begin{cases} \min\{x_1, \ldots, x_{M_s}\} < x_{u,\text{opt}} < \max\{x_1, \ldots, x_{M_s}\}, \\ \min\{y_1, \ldots, y_{M_s}\} < y_{u,\text{opt}} < \max\{y_1, \ldots, y_{M_s}\}. \end{cases} \tag{28}$$

Since the optimization problem has a large number of local maximum points, (28) helps us to limit the state and action space, and as a result, the convergence speed increases multiple times. In addition, according to (18g), the flying height of the UAV is limited as $h_{\min} \le z_u(t) \le h_{\max}$. Let us define the limited state space as

$$\mathcal{S} = \begin{cases} x_u(t) \in \left[\min\{x_1, \ldots, x_{M_s}\}, \max\{x_1, \ldots, x_{M_s}\}\right], \\ y_u(t) \in \left[\min\{y_1, \ldots, y_{M_s}\}, \max\{y_1, \ldots, y_{M_s}\}\right], \\ z_u(t) \in [h_{\min}, h_{\max}]. \end{cases} \tag{29}$$

Based on this, we modify the reward and action as follows:

$$r(t) = \begin{cases} \frac{-\ln(P_{\text{out}}(t))}{C_1 + a(t)}, & (s(t) + a(t)) \in \mathcal{S}, \\ 0, & (s(t) + a(t)) \notin \mathcal{S}, \end{cases} \tag{30}$$

$$a(t) = 0, \quad \text{if} \quad (s(t) + a(t)) \notin \mathcal{S}. \tag{31}$$

In (30), the term $a(t) + C_1$ is used to include the effects of the trajectory time in the reward. More precisely, if the action $a(t)$ is larger, it will definitely take more time. Also, the fixed term C_1 has been used so that the UAV does not receive unreasonably large rewards for small actions.

3) ALGORITHM

Our proposed trajectory method is summarized in Algorithm 1 and with schematic illustrated in Fig. 2(a). Algorithm 1 is provided for a dynamic network whose topology changes every T seconds, where $T \in \{T_{\min}, T_{\max}\}$ is a random variable. In Algorithm 1, the variables s_0'' and $s'(t')$ indicate the initial and final position of the UAV in each trajectory. The UAV flies from point s_0'' to point $s'(t')$ based on the trajectory obtained from Algorithm 1. The UAV stops at point $s'(t')$ until the network topology changes.

C. ADAPTIVE TD3-BASED METHOD

We show that with the geometrical analysis of the environment, the antenna patterns and transmitted power can be adjusted in such a way that the interference between the fronthaul links is reduced. To tackle this issue, as RL is a good approach for solving our considered dynamic optimization problem for trajectory, our solution is based on a hybrid technique by combining RL and geometrical analysis.

Proposition 1: If the OP of ith fronthaul link is limited by interference, then the optimal value of the antenna pattern is:

$$N_{ui,\text{opt}} = N_{\max}. \tag{32}$$

Proof: Based on the results of Appendix A, when the ith fronthaul link is limited by interference, its OP is proportional to

$$\mathbb{P}_{\text{out},i} \propto Q\left(c_1 - \frac{c_2}{N_{ui}^2}\right), \tag{33}$$

where $c_1 = \frac{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2}{2\sigma_\theta\sqrt{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2}}$, and $c_2 = \frac{\ln(\gamma_{\text{th}})}{2\sigma_\theta\sqrt{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2}}$. The term $(c_1 - \frac{c_2}{N_{ui}^2})$ is an ascending function of N_{ui}.

Algorithm 1 TD3-Based Trajectory Algorithm

```
Input: $N_{ui}$, $P_t$, $h_{\min}$, $h_{\max}$, $\alpha$, $b$, $\gamma$, $-\epsilon_{\min}$, $\epsilon_{\max}$, $s''_0$, $\bar{T}$
Output: Trajectory $\mathcal{B}_u(t)$
    Initialize all, critic networks $Q(s_t, a_t, \psi_1), Q(s_t, a_t, \psi_2)$ with $\psi_1, \psi_2$, actor network $\pi(s_t, \phi)$ with $\phi$
    Initialize target networks $\psi'_1 \leftarrow \psi_1$, $\psi'_2 \leftarrow \psi_2$, $\phi' \leftarrow \phi$
 1: Initialize environment and reset $S_i$ for $i = \{1, \ldots, M_s\}$.
 2: Initialize replay buffer
 3: Reset $s(1) = s''_0$. Generate random $T \in \{T_{\min}, T_{\max}\}$.
 4: for episode = 1 to max-number-episodes do
 5:     Initialize local buffer
 6:     Observe the initial state $s(t)$
 7:     for $n = 1$ to max-episode-steps do
 8:         Perform action $a(n) = \pi(s(n), \phi) + \epsilon$.
 9:         Observe reward $r(n)$ and the next state $s(n+1)$.
10:         Store the transition $(s(n), a(n), r(n), s(n+1))$ in replay buffer.
11:         Sample mini-batch from replay buffer.
12:         Update $\psi_1$ and $\psi_2$ by minimizing $J_{\psi_1}$ and $J_{\psi_2}$.
13:         if $d_{\mathrm{del}}$ mod $n$ then
14:             Update actor $\phi$ by minimizing (23).
15:             Update target networks: $\psi'_1 \leftarrow \tau\psi_1 + (1-\tau)\psi'_1$, $\psi'_2 \leftarrow \tau\psi_2 + (1-\tau)\psi'_2$, $\phi' \leftarrow \tau\phi + (1-\tau)\phi'$.
16:         end if
17:     end for
18: end for
19: Set $s'(0) = s''_0$, $a(0) = 0$, $v(0) = 1$, and $t = 0$.
20: for $n$ to max-number-episodes do
21:     Update $t = t + a(t)/v(t)$
22:     Get action $a(t) = \pi(s'(t), \phi)$
23:     Update $s'(t)$ based on $a(t)$ and store $s'(t)$ in $\mathcal{B}_u(t)$.
24:     Update $v(t)$ based on (29).
25: end for
26: Compute $P_{\mathrm{out}}$ for any elements of $\mathcal{B}_u(t)$ based on (26).
27: Find $s'(t')$ that has the least $P_{\mathrm{out}}$ in the range of $t < \bar{T}/2$.
28: Trajectory: Fly the UAV from $s'(0)$ to $s'(t')$ based on (1).
29: while $t < T$ do
30:     Stay UAV at $s'(t')$.
31: end while
32: Reset $s''_0 \leftarrow s'(t')$.
33: return  to line 1
```

Since $Q(x)$ is a descending function, therefore, in an interference-limited state, $\mathbb{P}_{\mathrm{out},i}$ is a decreasing function of N_{ui} and its optimal value is equal to the maximum value, i.e., $N_{ui,\mathrm{opt}} = N_{\max}$. ■

Based on the results of Proposition 1, when the performance is limited by interference, the best way is to increase N_{ui} and reduce the beamwidth. However, it should be noted that by increasing N_{ui}, the performance may change to the case limited by SNR. In this case, reducing the beamwidth is not necessarily the best option, because by reducing the beamwidth, the sensitivity of the system to the alignment error increases.

Algorithm 2 Adaptive TD3-Based Trajectory Algorithm

```
Input: $N_{\min}$, $N_{\max}$, $h_{\min}$, $h_{\max}$, $\alpha$, $b$, $\gamma$, $-\epsilon_{\min}$, $\epsilon_{\max}$, $s''_0$, $\bar{T}$
Output: Trajectory $\mathcal{B}_u(t)$, $\mathcal{N}'_u(t)$, $\mathcal{P}'_t(t)$
    Initialize all, critic networks $Q(s_t, a_t, \psi_1), Q(s_t, a_t, \psi_2)$ with $\psi_1, \psi_2$, actor network $\pi(s_t, \phi)$ with $\phi$
    Initialize target networks $\psi'_1 \leftarrow \psi_1$, $\psi'_2 \leftarrow \psi_2$, $\phi' \leftarrow \phi$
1: Do lines 1-8 of Algorithm 1.
2: Observe next state $s(n+1)$.
3: Compute adaptive values for $N_{ui}$ and $P_{t_i}$ by using the results of Propositions 1-3 for $i \in \{1, \ldots, M_s\}$.
4: Observe reward $r(n)$.
5: Do lines 10-23 of Algorithm 1.
6: Compute adaptive values for $N_{ui}(t)$ and $P_{t_i}(t)$ by using the results of Propositions 1-3 and store them in $\mathcal{N}'_u(t)$, and $\mathcal{P}'_t(t)$, respectively.
7: Do lines 24-32 of Algorithm 1.
8: return  to line 1
```

Proposition 2: If the OP of the F_i link is limited by SNR, the OP is a convex function of N_{ui} and its optimal value is obtained easily as:

$$N_{ui,\mathrm{opt}} = \min_{N_{ui}} \left| 1 + \left(\frac{c_1}{N_{ui}^2}\right)^{\frac{N_{ui}^2}{c_2}} \ln\left(\frac{c_1}{N_{ui}^2}\right) \right|, \tag{34}$$

where

$$\begin{cases} c_1 = \dfrac{\sigma_N^2 \gamma_{\mathrm{th}}}{P_{t_i} |h_{L_i}|^2 G_0(N_{si}) \pi}, \\ c_2 = \dfrac{(1.061)^2}{2\sigma_\Theta^2}. \end{cases} \tag{35}$$

Proof: Please refer to Appendix B. ■

Proposition 3: If the OP of the F_i link is limited by interference, the power changes have no effect on the system performance. In this case, in order to reduce the average transmitted power of the UAV, P_{t_i} can be reduced as follows:

$$P_{t_i} = \frac{\sigma_N^2 \gamma_{\mathrm{th}}}{|h_{L_i}|^2 G_0(N_{si}) G_0(N_{ui})} \left(\frac{P_{\mathrm{LoS}}(\theta_{\mathrm{ei}})}{\mathbb{J}_{2i}}\right)^{\frac{2\sigma_\Theta^2}{w_B^2(N_{ui})}}, \tag{36}$$

where

$$\mathbb{J}_{2i} = c_3\big(1 + \mathbb{P}''_{\mathrm{out},i} - P_{\mathrm{LoS}}(\theta_{\mathrm{ei}})\big). \tag{37}$$

Proof: Please refer to Appendix C. ■

Using the results of Propositions 1-3, we present an adaptive TD3-based algorithm for the trajectory problem that significantly improves the system performance. Our proposed adaptive method is summarized in Algorithm 2 and with schematic illustrated in Fig. 2(b).

D. POSITIONING

If the topology changes in the order of minutes, then the UAV has enough time to find the optimal position. Notice that our problem is formed by a large number of local maximum points for UAV position. However, since in the TD3

Algorithm 3 Positioning With Adaptive Pattern and Power Allocation

```
Input: $N_{\min}$, $N_{\max}$, $h_{\min}$, $h_{\max}$, $\alpha, b$, $\gamma, -\epsilon_{\min}, \epsilon_{\max}$, $m_{\text{ep}}$, $m'_{\text{ep}}$
Output: $B_{u,\text{opt}}$, $N_{ui,\text{opt}}$, $P_{t_i,\text{opt}}$ for $i \in \{1, \ldots, M_s\}$
    Initialize all, critic networks $Q(s_t, a_t, \psi_1), Q(s_t, a_t, \psi_2)$ with $\psi_1, \psi_2$, actor network $\pi(s_t, \phi)$ with $\phi$
    Initialize target networks $\psi'_1 \leftarrow \psi_1$, $\psi'_2 \leftarrow \psi_2$, $\phi' \leftarrow \phi$
Do lines 1 and 2 of Algorithm 1.
Initialize $r_{opt} = 0$
for $m = 1$ to $(m_{\text{ep}}/m'_{\text{ep}})$ do
    Generate random $s''_0$ and reset $s(1) = s''_0$.
    for episode = 1 to $m'_{\text{ep}}$ do
        Do lines 5-8 of Algorithm 1.
        Do lines 2-4 of Algorithm 2.
        if $r(n) > r_{\text{opt}}$ then
            Replace $r_{\text{opt}} = r(n)$, $B_{u,\text{opt}} = S(n+1)$, $N_{ui,\text{opt}} = N_{ui}$, $P_{t_i,\text{opt}} = P_{t_i}$ for $i \in \{1, \ldots, M_s\}$.
        end if
        Do lines 10-16 of Algorithm 1.
    end for
end for
```

algorithm, we follow the gradient to find the best parameters, we are guaranteed to converge on a local maximum (in most cases) or global maximum (best case). For the positioning approach, we again exploit a TD3-based algorithm, which efficiently solves the non-convex positioning optimization problem in (19). To approach the global maximum, it is enough to run TD3 algorithm in a large number. Unlike in the trajectory case, the trajectory time here is not important and thus, we modify the reward as:

$$r(t) = \begin{cases} -\ln(P_{\text{out}}(t)), & (s(t) + a(t)) \in \mathcal{S}, \\ 0, & (s(t) + a(t)) \notin \mathcal{S}. \end{cases} \tag{38}$$

Moreover, the starting point is not important and we can start the algorithm from any arbitrary point. In addition, the simulation results show that starting from different random points makes a better exploration in the state space and approaches the global maximum point faster. However, the different number of starting points cannot be chosen too much and depends on the number of episodes. This is due to this point that the combination of exploration and exploitation in TD3 algorithm helps us to reach the maximum points. Our proposed adaptive method for UAV positioning is summarized in Algorithm 3. Parameters m_{ep} and m'_{ep} in the input of Algorithm 3 indicate the total number of episodes and the number of episodes for each random initial position, respectively. As we show in the simulations, the choice of the starting points is very important in the faster convergence of Algorithm 3. Therefore, in the sequel, by investigating critical links, we can find a relative understanding of the approximate position of the global optimal point. Then we use the approximate points as the starting points and it is observed that the algorithm converges several times faster.

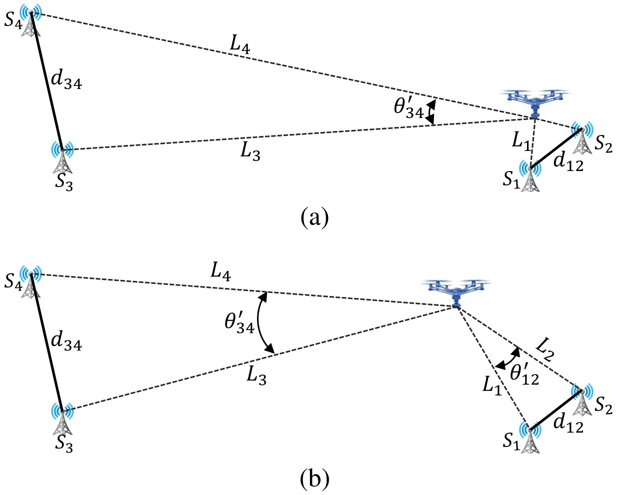


FIGURE 4. A graphical example of how close SBSs affect the optimal position of the UAV. (a) The close proximity of the UAV to S_1 and S_2 decreases spatial angle θ_{34}.
(b) The UAV readjusts its position based on the position of S_is and parameters d_{12} and d_{34} in such a way that a balance is created in the spatial angles θ_{12} and θ_{34}.

1) CRITICAL FRONTHAUL LINKS

For the considered system, the fronthaul links generally meet the condition of the requested OP, except in the several following critical cases. First, having two SBSs close to each other causes a small spatial angle between their fronthaul links. Without loss of generality, we specify the two SBSs with the smallest distance to each other with S_1 and S_2. Also, $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ is the distance between S_i and S_j. The length of F_i link is obtained as

$$L_i = \sqrt{(u_x - x_i)^2 + \left(u_y - y_i\right)^2 + h_u^2}, \tag{39}$$

where h_u is the instantaneous height of the UAV. The spatial angle between F_i and F_j links is obtained as

$$\theta'_{12} = \cos^{-1}\left(\frac{L_1^2 + L_2^2 - d_{12}^2}{2L_1L_2}\right). \tag{40}$$

For the critical condition, we know that d_{12} is very small. Therefore, $d_{12} << L_1$ and then we have $L_1 \simeq L_2$. Using this, θ'_{12} can be well approximated as

$$\theta'_{12} \simeq \cos^{-1}\left(1 - \frac{d_{12}^2}{2L_1L_2}\right). \tag{41}$$

According to (41), θ'_{12} increases with the decrease of L_1 and L_2. Therefore, at the beginning of the trajectory, the UAV first decides to fly towards the critical SBSs S_1 and S_2.

However, by getting too close to the critical points S_1 and S_2, it may cause other spatial angles θ'_{ij} to become in a critical situation. In order to get a better understanding, a graphical example is provided in Fig. 4(a). As we see, although $d_{34} > d_{12}$, since the UAV being too close to S_1 and S_2, we have $\theta_{34} < \theta_{12}$. In this case, using (40), the UAV corrects its position in such a way that the following relationship approximately created:

$$\frac{L_1^2 + L_2^2 - d_{12}^2}{2L_1L_2} \simeq \frac{L_3^2 + L_4^2 - d_{34}^2}{2L_3L_4}. \tag{42}$$

There are two important points related to (42). First, (42) has several solutions, which are a function of the UAV's height. If the only goal is to reduce the spatial angle between fronthaul links at critical points, the OP decreases by reducing the UAV's height. Moreover, reducing the UAV's height decreases the link length and thus, it reduces the channel propagation loss. However, a decrease in the UAV's height decreases the elevation angle. Therefore, based on (5) and (8), the SBSs that are at a farther distance from the UAV become in critical conditions. Second, the UAV's position obtained form (42) is not the global optimal point. As mentioned, the global optimal point is a function of many correlated parameters, which cannot be accurately calculated using geometrical analysis. However, the results of (42) can be used as a starting point in Algorithm 3. Using these results obtained from the instantaneous geometry of the network, we modify the starting point of Algorithm 3. For different values of the UAV's height, we obtain the position of UAV from (42), and then, we use it as the starting point in Algorithm 3. The simulation results show that the use of start points obtained from (42) increases the convergence speed of Algorithm 3, significantly.

E. TRAJECTORY/POSITIONING UNDER ACTUAL ENVIRONMENT

In the previous sections, according to the instantaneous θ_{ei}s, we calculated the reward (based on outage probability) by considering the LoS probability. However, we can add more complexity to the trajectory/positioning problem by assuming that the UAV can quickly determine its LoS status. Let $f_{B_u,i}$ determine the LoS status of position B_u relative to the S_i. Therefore, the UAV is either in the LoS of S_i wherein $f_{B_u,i} = 1$ or in the NLoS state wherein $f_{B_u,i} = 0$. Therefore, under this condition, (26) is modified as:

$$\begin{aligned}\mathbb{P}_{\mathrm{out},i} \simeq\ & Q\left(\frac{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2 - w_B^2(N_{ui})\ln(\gamma_{\mathrm{th}})}{2\sigma_\theta\sqrt{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2}}\right) f_{B_u,i} \\ & + \left(\frac{\sigma_N^2\gamma_{\mathrm{th}}}{P_{t_i}|h_{L_i}|^2 G_0(N_{si})G_0(N_{ui})}\right)^{\frac{w_B^2(N_{ui})}{2\sigma_\Theta^2}} \\ & \times U\left(1 - \frac{\sigma_N^2\gamma_{\mathrm{th}}}{P_{t_i}|h_{L_i}|^2 G_0(N_{si})G_0(N_{ui})}\right) f_{B_u,i} \\ & + \mathbb{P}_{\mathrm{out,NLoS}}\left(1 - f_{B_u,i}\right),\end{aligned} \tag{43}$$

where $\mathbb{P}_{\mathrm{out,NLoS}}$ is the outage probability in the NLoS state. For the NLoS path, the received signal is caused by the refraction or reflection of the signal. Due to the high attenuation of refraction or reflection of the signal at high frequencies, the received signal is strongly weakened in NLoS state. More importantly, in NLoS state, for directional antennas, the received signal is caused by the reflection of the signal sent from the antenna's side lobes, which are placed on the side lobes of the receiving antenna. As much as the main gain of the antenna increases, the gain of the side lobes decreases. Therefore, in the NLoS state, the outage probability tends to one. In this case, based on (25), we have:

$$r_t = -\ln\left(P_{\mathrm{out,NLoS}} \simeq 1\right) \simeq 0 \tag{44}$$

To help the UAV get out of NLoS state faster, we modify the reward defined in (30) as:

$$r(t) = \begin{cases} \frac{-\ln(P_{\mathrm{out}}(t))}{C_1 + a(t)}, & (s(t) + a(t)) \in \mathcal{S},\ f_{B_u,i} = 1, \\ -1, & (s(t) + a(t)) \in \mathcal{S},\ f_{B_u,i} = 0, \\ 0, & (s(t) + a(t)) \notin \mathcal{S}. \end{cases} \tag{45}$$

Also, by increasing the height, the probability of being in the LoS state increases, and thus, we modify the action along the z axis as:

$$a_z(t) = (z_u(t) + h_{\max})/2, \quad \text{if} \quad f_{B_u,i} = 0. \tag{46}$$

Finally, in order to prepare Algorithms 1–3 under actual environment, (43) and (45) should be used to calculate rewards, and (46) should be used to perform actions along the z axis.

V. SIMULATION RESULTS AND PERFORMANCE ANALYSIS

A. TRAJECTORY

For simulations, we consider that the UAV covers an area of 150×150 m^2. The UAV is equipped with $M_u = 8$ antennas at frequency $fc = 140$ GHz. One of the practical problems of using THz frequencies on UAVs is the power amplifier whose dimensions is large [38]. Therefore, we considered the maximum transmitted power of 10 mW for each antenna, which is practically possible for installation on a UAV. To compensate the low transmitted power, we have used the array antennas on the UAV. Each square array antenna includes of $N_{ui} \times N_{ui} = 20 \times 20$ elements with equal spacing of $d = \lambda/2$ between elements. Therefore, the effective size of each array antenna is $A_{\mathrm{eff}} \simeq \frac{20\times c}{f_c} = 4.3$ cm^2 [26], which has acceptable dimensions for installation on the UAV. We have also assumed that the SBSs are distributed with a uniform random distribution and their topology changes every T second. Parameter $T \in \{T_{\min}, T_{\max}\}$ is also a random variable, where $T_{\min} = 20$, and $T_{\max} = 35$. The maximum speed of the UAV is $v_{\max} = 8$ m/s, and its acceleration is $v' = 4$ m/s^2. The UAV has a flight height limit of $h_{\min} = 30$ m, and $h_{\max} = 130$ m. Also, the intensity of the UAV's vibrations is considered as $\sigma_\theta = 2°$.

MATLAB software is used to implement the algorithms, define the 3D environment, adjust antenna patterns, and compute rewards. Moreover, to train TD3 agents we use the Reinforcement Learning Designer app in [39]. For the machine learning configurations, all neural networks are initialized with the same parameters: each has two fully-connected hidden layers with size=256 × 128 neurons. Adam is used as the optimizer of both critic and actor

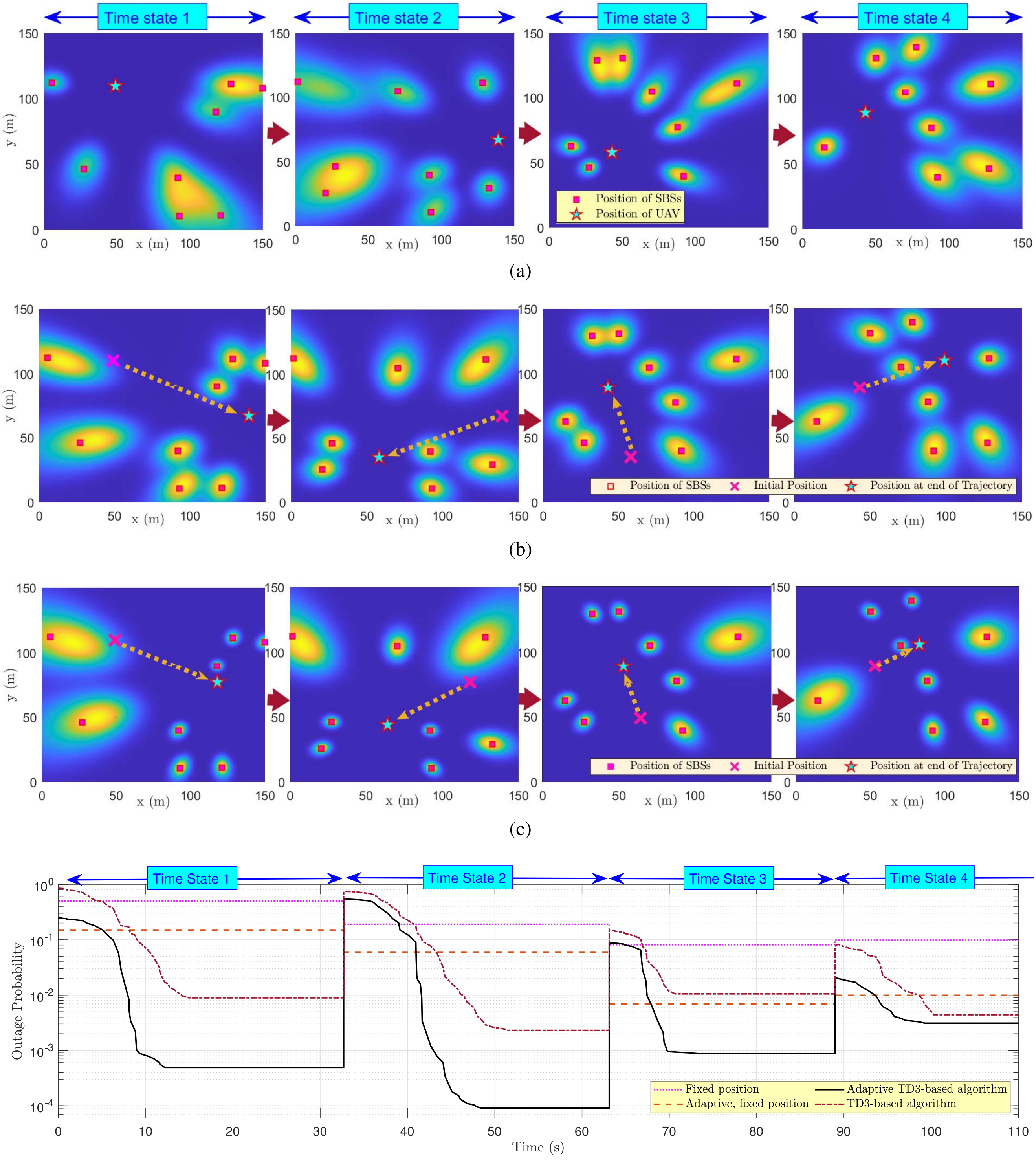


FIGURE 5. Studying the UAV trajectory in 4 consecutive time states. These figures show the power distribution (a) at the starting point of each time state, (b) at the final point of the trajectory obtained from Algorithms 1, and (c) at the final point of the trajectory obtained from Algorithms 2. (d) The OP of the considered system for all four consecutive time states is plotted versus time. To show the importance of trajectory, the results of Algorithms 1 and 2 are compared with a stationary system.

networks. The hyper-parameters are set as follows: the learning rate of both the actor and critic networks are 10^{-4}, the discount factor $\gamma = 0.9$, the mini-batch size=32, replay buffer size=1000, maximum number of episodes 200, the maximum steps per episode is 20 and the Poylal averaging factor $\tau = 0.01$. Moreover, during the learning process, sometimes Q-functions begins to dramatically overestimate Q-values, which then leads to the policy breaking. To avoid this issue, we run the algorithms four times in parallel and consider the trajectory that has the least OP.

In Fig. 5, we investigate the effect of the trajectory on the OP of the considered dynamic system. To this end, we have considered the UAV trajectory in 4 consecutive time states. In each time state, a number of SBSs are randomly disconnected and a number of new SBSs with random positions are connected to the UAV. With any change in network topology, the last position of the UAV in the previous time state becomes the starting point of the trajectory in the next time state. Figures 5(a)-5(c) show the scaled power distribution of the UAV. In particular, Fig. 5(a) shows the power distribution at the starting point of each time state. Figures 5(b) and 5(c) show the power distribution at the final point of the trajectory obtained from Algorithms 1 and 2, respectively. The results of Fig. 5(a) clearly show that by changing the topology of the network, the interference between SBSs increases at the start point, especially for SBSs that are close to each other. In this case, we first use Algorithm 1 for trajectory, which navigates the UAV to an optimal point in the 3D space. By comparing Figs. 5(a) and 5(b), it can be seen that the UAV manages the power distribution by modifying its position in such a way that the interference between the SBSs is reduced. The results of Fig. 5(c) are for the trajectory obtained from Algorithm 2, in which the UAV, in addition to modifying the position, corrects its antenna pattern. The results of Fig. 5(c) show well how the combination of trajectory and antenna pattern has further reduced the interference between distributed SBSs. The UAV uses antenna patterns with very small beamwidth to reduce the interference between SBSs that are close to each other. However, the OP of these links becomes more sensitive to antenna misalignment due to the UAV's vibrations. On the other hand, for the rest of the SBSs that have a sufficient distance from the other SBSs, a larger antenna beamwidth is selected to achieve a trade-off between the antenna gain and the pointing errors.

It should be noted that in Figs. 5(a)-5(c), the position of the UAV is shown on the $x-y$ plane. In practice, the UAV, in addition to its position on the $x-y$ plane, also modifies its height, which cannot be displayed in these figures. The results of the figures provide an overview of how UAV manages power distribution between SBSs using trajectory. To find more information, for all four consecutive time states, the OP of the considered system is also plotted in Fig. 5(d). The results of this figure give us more information about the time of the UAV trajectory from the starting point to the final point. In Fig. 5(d), to show the importance of the trajectory, we have compared its performance with the stationary systems. For a stationary system, the best position is definitely in the center because it provides larger spatial angle between SBSs on average. As the results of Fig. 5(d) show, the stationary system generally performs better at the start point. But the UAV corrects its position in a short time and significantly improve the OP. It should be noted that the results show the OP of the worst fronthaul link (critical link) and the rest of the links have lower OPs.

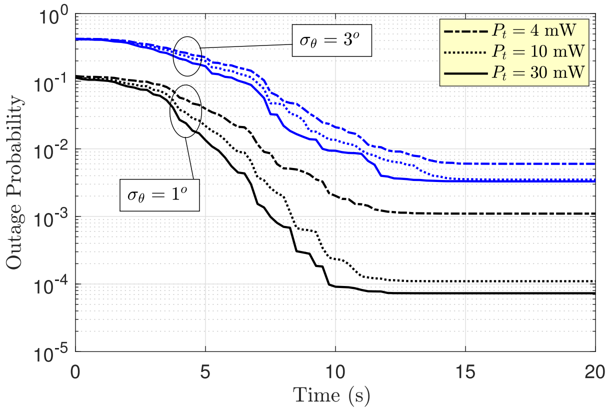


FIGURE 6. OP versus t for three different values of P_t and two different values of σ_θ.

In the following, we examine the effect of the transmitted power and the intensity of UAV instability on the system performance along the trajectory. To this end, in Fig. 6, for a random positioning of SBSs and taking into account a random start point, the OP of the considered system is depicted for three different values of P_t, as well as two different values of UAV instability, $\sigma_\theta = 1°$ and $3°$. As the simulations show, for $\sigma_\theta = 1°$, by increasing P_t from 4 to 10 mW, the system performance improves significantly. However, by increasing P_t from 10 to 30 mW, the performance of the system is slightly improved, which indicates that the system is limited by interference. In this case, increasing the power simultaneously increases the interference. For $\sigma_\theta = 3°$, the probability of interference increases, and as a result, it can be seen that increasing P_t has a negligible effect on the system performance. The general result is that in the interference-limited mode, and where two SBSs are close to each other, increasing P_t cannot improve the system performance.

B. POSITIONING

Next, we examine the issue of positioning. To this end, in Table 1, we investigate the performance of the algorithms provided for positioning. For the benchmark, we have used exhaustive search algorithm. For the exhaustive search, we divided the 3D space into discrete parts with an accuracy of 50 cm. The total search space is 200×300×300. In the whole discrete search space, we find the optimal position of the UAV with the least OP. Algorithm 3 was first proposed for positioning. Then, by studying the statistical geometry of the environment, we modified the starting point of Algorithm 3 according to the distribution of SBSs. We have implemented the algorithms for different values of time steps. The number of time steps is equal to the number of episodes multiplied by the steps per episode. The steps per episode is considered equal to 10. The simulation results clearly show that using the starting points obtained from the network geometry helps the machine learning algorithm to converge faster. As it can

TABLE 1. Comparison of the proposed algorithms for positioning with exhaustive search results.

Algorithm	Adaptive						Non-adaptive					
	Time steps	OP	x_u	y_u	z_u	$\overline{P}_t$ mW	States	OP	x_u	y_u	z_u	$\overline{P}_t$ mW
Exhaustive search	$300 \times 300 \times 200$	1.3×10^{-4}	110.5	40.5	75	31.3	$300 \times 300 \times 200$	2.5×10^{-3}	117.5	31.5	66	80
TD3 with selected start points	500	1.4×10^{-3}	98.3	58.2	71.7	28.7	500	9.8×10^{-3}	112.1	42.8	45.8	80
	1000	3.5×10^{-4}	96.4	47.1	80.2	29.2	1000	6.2×10^{-3}	110.8	45.4	51.2	80
	5000	2.3×10^{-4}	104.6	47.3	83.1	30.6	5000	3.6×10^{-3}	122.5	32.2	57.2	80
	10000	2.1×10^{-4}	108.9	37.2	77.6	31.5	10000	2.7×10^{-3}	118.4	35.7	67.1	80
	50000	1.3×10^{-4}	110.3	40.4	74.8	31.2	50000	2.5×10^{-3}	117.6	31.3	66.3	80
TD3 with random start points	500	5.2×10^{-2}	100.2	78.6	56.1	26.1	500	2.1×10^{-1}	67.4	63.4	43.1	80
	1000	1.1×10^{-2}	88.7	74.2	63.2	28.4	1000	5.6×10^{-2}	79.7	57.2	47.1	80
	5000	2.3×10^{-3}	101.8	53.9	81.7	29.5	5000	8.8×10^{-3}	93.3	47.1	82.6	80
	10000	4.5×10^{-4}	117.2	46.6	75.6	30.4	10000	5.2×10^{-3}	123.2	32.4	73.2	80
	50000	1.4×10^{-4}	111.1	40.1	74.8	31.1	50000	2.5×10^{-3}	117.7	33.1	68.2	80

be seen, with only 1000 time steps, it achieves the OP close to the lowest OP obtained from exhaustive search.

The results of Table 1 include two adaptive and non-adaptive parts. The adaptive algorithm refers to an algorithm that along with positioning, selects the optimal values for antenna patterns and powers. By comparing the results of two parts, the adaptive algorithm performs much better than non-adaptive algorithm. It should be noted that the computational complexity of the adaptive algorithm is almost the same as that of the non-adaptive algorithm, because by using the results of Propositions 1-3, we determine the optimal values according to each UAV position. Moreover, from a practical point of view, almost no special complexity is added to the adaptive algorithm compared to the non-adaptive algorithm. In particular, both adaptive and non-adaptive algorithms have the same maximum transmitted power. Only the adaptive algorithm reduces the transmitted power in some times when the system is limited by interference, because it was shown that the power causes the simultaneous amplification of the main signal and the interference. Moreover, in order to choose the optimal pattern, it is enough to make the antenna elements active or inactive, without the need for heavy processing to beam shaping. The dimensions of the considered array antenna were also discussed at the THz frequency, which is in the order of 4 cm.

C. 3D ENVIRONMENT

We can use different methods such as ray tracing to model LoS in an environment with real obstacles. Since the computational volume and the processing speed are key parameters in our analysis, in order to model LoS in a 3D environment with real obstacles, a method based on quantization of the 3D environment is provided in Appendix D, which can be implemented with high accuracy and low computational volume.

To simulate the obstacles, we assume that in an area 150 × 150 m^2, 10 buildings with an average height of 15 meters and a length and width of 8 meters are randomly distributed. The SBSs are randomly distributed similar to the previous section. The UAV also starts working in an initial random location. The 3D image of the distribution of obstacles (buildings) is shown in Fig. 7(a). According to the modified Algorithms 1 and 2, the UAV starts the trajectory. Algorithm 1 is for the case with a fixed pattern and Algorithm 2 is for the case with an adaptive pattern. The outage probability of links during the trajectory is plotted in Fig. 7(b) for both algorithms. As we can see, for both algorithms, at the start times, the UAV has a maximum outage probability equal to 1. During these start times, at least one of the SBSs is in the NLoS state. At the moment $t = 5.9$ s, the UAV solves the problem related to NLoS of all SBSs and at this moment the outage probability decreases, significantly. For the rest of the trajectory, the UAV acts the same as before to control the interference and the distance between the SBSs to minimize the maximum outage probability.

VI. CONCLUDING REMARKS AND FUTURE DIRECTIONS

In this work, we consider a general dynamic UAV-based HetNet taking into account the random positioning of SBSs, spatial angles between THz links, real antenna pattern, and UAV's vibrations in the 3D space. Using directional THz antennas, the main goal of this work was spatial frequency reuse to increase the capacity of a dynamic UAV-based HetNet, while improving the quality of service of each link (we considered OP as the service quality metric). Then, using geometrical analysis and deep RL method, we proposed several algorithms to learn the optimal trajectory/positioning to minimize the maximum OP of directional THz links. In the presence of UAV's vibrations, considering the importance of antenna pattern in interference management, we proposed an algorithm that selects an optimal pattern during the trajectory. We then modified our proposed algorithms for an environment with 3D obstacles and observed that by using the proposed algorithms, the UAV learns well during the trajectory to first place all the SBSs in the LoS state. Then, with the optimal antenna pattern selection, the UAV finds the optimal trajectory to reduce the interference between the SBSs that are close to each other.

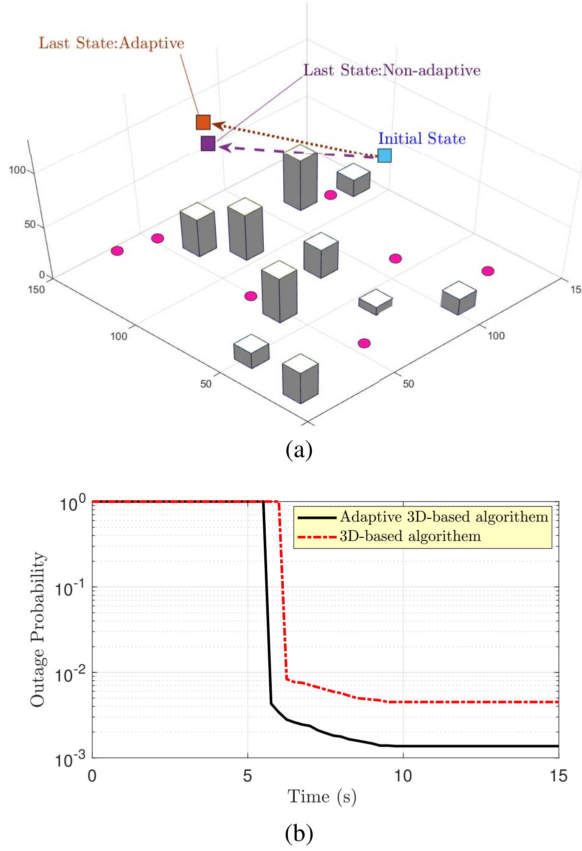


FIGURE 7. (a) 3D illustration of obstacles and random positioning of SBSs. (b) Comparison of the outage probability of the considered system during the trajectory in the 3D environment depicted in Fig. 7(a).

We envision several future research directions for this work. For instance, in this work, we tried to cover an area with a UAV. However, for wider areas, it is better to use several UAVs, in which case the trajectory problem becomes more complicated. In addition, for the conventional network with multiple UAVs using omnidirectional antennas, usually, the ground nodes are connected to the UAVs with the shortest distance. However, for the considered dynamic network equipped with directional antennas, the performance is mostly affected by the spatial angle between the nodes and the LoS status. Therefore, for a network with several UAVs, novel clustering methods must be adopted for connecting the ground nodes to the UAVs. Moreover, the main problem of the considered dynamic network is that when the ground nodes are very close to each other, high interference is generated. In order to manage interference, one can use new techniques such as non-orthogonal multiple access technique to manage interference during the trajectory.

APPENDIX A

By studying the behavior of the end-to-end SINR provided in (9) under different conditions, we provide an approximate closed-form expression of the OP, which is close to the exact OP.

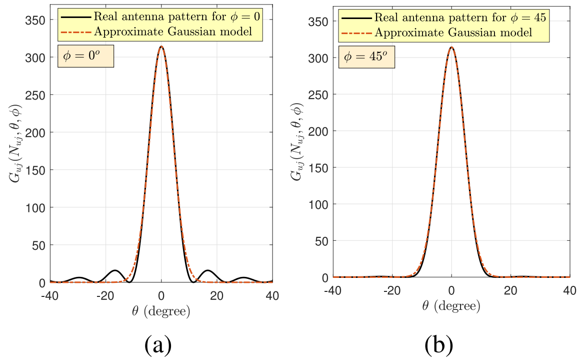


FIGURE 8. Showing the effect of the Azimuth angle ϕ on the side-lobe of an $N_{uj} \times N_{uj} = 10 \times 10$ array antenna pattern.

1) LIMITED BY SNR

According to results of [10], for larger spatial angles between fronthaul links, the side-lobes of the antenna pattern are the main cause of the interference. In these conditions, it was also shown in [10] that only by rotating the antenna pattern in the direction of the azimuth angle, the interference between fronthaul links can be significantly reduced to below the noise level. Therefore, to reduce the interference, it is enough to set the azimuth angle between the antenna patterns close to 45 degrees. To better understand this point, in Fig. 8, the real antenna pattern is drawn for two cases with $\phi = 0$ and $\phi = 45°$. As can be seen, for $\phi = 45°$, the interference effect of the side-lobes can be easily ignored. In this case, the antenna pattern is well approximated with the Gaussian main-lobe pattern as follows [40]:

$$G_{uj}(N_{ui}, \theta) = G_0(N_{ui}) \times \exp\left(-\frac{\left(\tan^{-1}\left(\sqrt{\tan^2(\theta_x) + \tan^2(\theta_y)}\right)\right)^2}{w_B^2(N_{ui})}\right), \quad (47)$$

where $w_B(N_{ui}) = \frac{1.061}{N_{ui}}$ is the angular beamwidth (called the beam divergence). According to the mentioned points, assuming $-40° < \phi < 50°$, for larger spatial angles between fronthaul links, (9) is simplified as follows:

$$\gamma_i \simeq \frac{P_{t_i} \alpha_{L_i} |h_{L_i}|^2 G_0^2(N_{si}) G_0^2(N_{ui})}{\sigma_N^2} \times \exp\left(-\frac{\left(\tan^{-1}\left(\sqrt{\tan^2(\Theta_x) + \tan^2(\Theta_y)}\right)\right)^2}{w_B^2(N_{ui})}\right). \quad (48)$$

In practice, the UAV's fluctuations $\Theta = [\Theta_x, \Theta_y]$ is less than 5°, in which (48) is simplified as follows:

$$\gamma_i \simeq \frac{P_{t_i} \alpha_{L_i} |h_{L_i}|^2 G_0^2(N_{si}) G_0^2(N_{ui})}{\sigma_N^2} e^{-\frac{\Theta_x^2 + \Theta_y^2}{w_B^2(N_{ui})}} \quad (49)$$

Substituting (49) in (17), we obtain

$$\mathbb{P}'_{\text{out},i} = \text{Prob}\left[\frac{\Theta_x^2+\Theta_y^2}{w_B^2(N_{ui})} > \ln\left(\frac{P_{t_i}\alpha_L|h_{L_i}|^2 G_0(N_{si})G_0(N_{ui})}{\sigma_N^2\gamma_{\text{th}}}\right)\right]. \quad (50)$$

Based on (7) and (50) and using [41], after some manipulations, $\mathbb{P}'_{\text{out},i}$ is derived as

$$\mathbb{P}'_{\text{out},i} = (1 - P_{\text{LoS}}(\theta_{ei})) + P_{\text{LoS}}(\theta_{ei})\left(\frac{\sigma_N^2\gamma_{\text{th}}}{P_{t_i}|h_{L_i}|^2 G_0(N_{si})G_0(N_{ui})}\right)^{\frac{w_B^2(N_{ui})}{2\sigma_\Theta^2}} \times U\left(1 - \frac{\sigma_N^2\gamma_{\text{th}}}{P_{t_i}|h_{L_i}|^2 G_0(N_{si})G_0(N_{ui})}\right). \quad (51)$$

2) LIMITED BY SIR

Although by adjusting the azimuth angle, the interference caused by the side-lobes can be reduced, for smaller spatial angles between the fronthaul links, the interference is caused by the main lobe of the antennas. In this case, SINR is limited by interference and (9) is simplified as:

$$\gamma_i \simeq \frac{G_0(N_{ui})\exp\left(-\frac{\left(\Theta_x^2+\Theta_y^2\right)^2}{w_B^2(N_{ui})}\right)}{G_0\left(N_{uj}\right)\exp\left(-\frac{\left(\theta_{x_{ij}}+\Theta_x\right)^2+\left(\theta_{y_{ij}}+\Theta_y\right)^2}{w_B^2(N_{uj})}\right)} \quad (52)$$

where i and j are the two fronthaul links that have the smallest spatial angle to each other. The important point is that the average interference of link i on link j is equal to the average interference of link j on link i. Due to symmetry, the optimal pattern for both links i and j should be equal. Therefore, (52) is simplified as follows:

$$\gamma_i \simeq \exp\left(\frac{\theta_{x_{ij}}^2 + 2\theta_{x_{ij}}\Theta_x + \theta_{y_{ij}}^2 + 2\theta_{y_{ij}}\Theta_y}{w_B^2(N_{ui})}\right). \quad (53)$$

Using (17) and (53), after some manipulations, the outage probability of link i when it is limited by interference is derived as

$$\mathbb{P}''_{\text{out},i} = Q\left(\frac{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2 - w_B^2(N_{ui})\ln(\gamma_{\text{th}})}{2\sigma_\theta\sqrt{\theta_{x_{ij}}^2 + \theta_{y_{ij}}^2}}\right). \quad (54)$$

Finally, using (51) and (54), the outage probability of link i is derived in (26).

APPENDIX B

Based on the results of Appendix A, when the ith fronthaul link is limited by SNR, its OP is proportional to

$$\mathbb{P}'_{\text{out},i} \propto \mathbb{J}_1(N_{ui}) = \left(\frac{c_1}{N_{ui}^2}\right)^{\frac{c_2}{N_{ui}^2}}. \quad (55)$$

$c_1 = \frac{\sigma_N^2\gamma_{\text{th}}}{P_{t_i}|h_{L_i}|^2 G_0(N_{si})\pi}$, and $c_2 = \frac{(1.061)^2}{2\sigma_\Theta^2}$. We can rewrite relation (55) as follows:

$$\mathbb{J}_1(N_{ui}) = \exp\left(\frac{c_2}{N_{ui}^2}\ln\left(\frac{c_1}{N_{ui}^2}\right)\right). \quad (56)$$

By taking the derivative of (56), we get:

$$\mathbb{J}'_1(N_{ui}) = -\left(\frac{c_1}{N_{ui}^2}\right)^{\frac{c_2}{N_{ui}^2}}\left(\frac{2c_2}{N_{ui}^3}\right)\ln\left(\frac{c_1}{N_{ui}^2}\right) - \frac{2c_2}{N_{ui}^3}. \quad (57)$$

Now, by deriving again from (57), we have:

$$\mathbb{J}''_1(N_{ui}) = \underbrace{-\frac{4c_2}{N_{ui}^4}\ln(\mathbb{J}_1(N_{ui}))}_{\text{Term}_1>0} + \underbrace{\frac{2}{N_{ui}}\mathbb{J}_1(N_{ui}) + \frac{6c_2}{N_{ui}^4}}_{\text{Term}_2>0}$$
$$\frac{2\mathbb{J}_1(N_{ui})}{N_{ui}^2}\underbrace{\left[2\ln^2(\mathbb{J}_1(N_{ui})) + 3\ln(\mathbb{J}_1(N_{ui}))\right]}_{\text{Term}_3>0\ \text{for}\ \mathbb{J}_1(N_{ui})<0.22}. \quad (58)$$

As can be seen, (58) consists of the sum of three terms. The first and second terms of which are always positive. The third term also guarantees that it is always positive for OP less than 0.22, which is almost always true, and thus, OP is a convex function of N_{ui}. Now, by setting (57) equal to zero, the optimal value for N_{ui} is obtain by solving following equation:

$$\mathbb{J}'_2 = 1 + \left(\frac{c_1}{N_{ui}^2}\right)^{\frac{c_2}{N_{ui}^2}}\ln\left(\frac{c_1}{N_{ui}^2}\right) = 0 \quad (59)$$

Note that the optimal N_{ui} obtained from (59) can be any positive real number while N_{ui} is an integer. Therefore, the obtained optimal N_{ui} from (59) is approximated to the nearest integer number which is the optimal value. On the other hand, we obtained that the OP is a convex function of N_{ui}. Using these points, the optimal value for N_{ui} is equal to the integer value where the sign of $\mathbb{J}'_2$ changes from negative to positive, which occurs only for one integer number. Using these points and based on (59), we can easily obtain the optimal N_{ui} as (34).

APPENDIX C

When the performance of the F_i link is limited by the interference of the F_j link, due to the instantaneous position of the UAV, the instantaneous value of the spatial angle θ_{ij} is small. Based on the results of Appendix A, (9) is simplified as follows:

$$\gamma_i \simeq \frac{P_{t_i}G_0(N_{ui})\exp\left(-\frac{\left(\Theta_x^2+\Theta_y^2\right)^2}{w_B^2(N_{ui})}\right)}{P_{t_j}G_0\left(N_{uj}\right)\exp\left(-\frac{\left(\theta_{x_{ij}}+\Theta_x\right)^2+\left(\theta_{y_{ij}}+\Theta_y\right)^2}{w_B^2(N_{uj})}\right)}. \quad (60)$$

Using the fact that $\theta_{ij} = \theta_{ji}$, and according to the symmetry, the optimal value for P_{t_i} is equal to P_{t_j}. Therefore, based on (60), by increasing P_{t_i}, interference increases as well. As a result, when the performance is limited by interference, power changes have no effect on the performance. In this case, P_{t_i} can be reduced to decrease the average transmitted power of the UAV. However, a large reduction of P_{t_i} can reduce SNR. Therefore, P_{t_i} can be reduced so that SIR is not less than SNR. Based on this point, by using (26), the optimal value for P_{t_i} is obtained as (36).

APPENDIX D

For the 3D simulation of obstacles, we first quantize each given environment with d_b precision and convert it into a 3D matrix $\mathbf{K}_u \in \mathbb{R}^{3\times 3}$, where each element of the matrix $k_{m,n,p}$ represents a small cube with d_b^3 volume. If there are obstacles in $k_{i,j,n}$, then $k_{i,j,n} = 1$, and if there are no obstacles, $k_{m,n,p} = 0$. Also, for each UAV position, the equation of the connecting line between S_i and the instantaneous UAV position is formulated as:

$$\frac{x - x_i}{x_u(t) - x_i} = \frac{y - y_i}{y_u(t) - y_i} = \frac{z}{z_u(t)}. \tag{61}$$

Then we define the matrix $\mathbf{K}_i' \in \mathbb{R}^{3\times 3}$ so that the dimension of $\mathbf{K}_i'$ is equal to $\mathbf{K}_u$. Then, we adjust the elements of $\mathbf{K}_i'$ in such a way that $k'_{m,n,p} = 1$ when the equation of line (61) is located at that element and the rest of the elements are set as $k'_{m,n,p} = 0$. Now, we obtain matrix $\mathbf{K} \in \mathbb{R}^{3\times 3}$ as

$$\mathbf{K} = \mathbf{K}_i \odot \mathbf{K}_u, \tag{62}$$

where $\odot$ represents the Hadamard (element-wise) product. If all the elements of $\mathbf{K}$ are equal to zero, this means there is no interference between the obstacles and the connecting line defined in (61), and thus, the UAV is in the LoS state of S_i. Otherwise, it is in the NLoS state. In this method, choosing an optimal value for d_b is very important because a small d_b should be chosen to achieve high accuracy. On the other hand, as d_b is chosen to be small, the dimensions of $\mathbf{K}_u$ and $\mathbf{K}_i$ increase, which leads to an increase in computational load.

ACKNOWLEDGMENT

The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] M. T. Dabiri and M. Hasna, "UAV trajectory optimization for directional THz links using deep reinforcement learning," in *Proc. IEEE 97th Veh. Technol. Conf. (VTC)*, 2023, pp. 1–5.

[2] M. Alzenad, M. Z. Shakir, H. Yanikomeroglu, and M.-S. Alouini, "FSO-based vertical backhaul/fronthaul framework for 5G+ wireless networks," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 218–224, Jan. 2018.

[3] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2109–2121, Mar. 2018.

[4] C. Shen, T.-H. Chang, J. Gong, Y. Zeng, and R. Zhang, "Multi-UAV interference coordination via joint trajectory and power control," *IEEE Trans. Signal Process.*, vol. 68, pp. 843–858, 2020.

[5] G. Zhang, Q. Wu, M. Cui, and R. Zhang, "Securing UAV communications via joint trajectory and power control," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1376–1389, Feb. 2019.

[6] Z. Ji, W. Yang, X. Guan, X. Zhao, G. Li, and Q. Wu, "Trajectory and transmit power optimization for IRS-assisted UAV communication under malicious jamming," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11262–11266, Oct. 2022.

[7] X. Liu, M. Chen, S. Wang, W. Saad, and C. Yin, "Trajectory design for energy harvesting UAV networks: A foraging approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2020, pp. 1–6.

[8] K. Meng, Q. Wu, S. Ma, W. Chen, and T. Q. S. Quek, "UAV trajectory and beamforming optimization for integrated periodic sensing and communication," *IEEE Wireless Commun. Let.*, vol. 11, no. 6, pp. 1211–1215, Jun. 2022.

[9] X. Lu, M. Salehi, M. Haenggi, E. Hossain, and H. Jiang, "Stochastic geometry analysis of spatial-temporal performance in wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2753–2801, 4th Quart., 2021.

[10] M. T. Dabiri, M. Hasna, and W. Saad, "Downlink interference analysis of UAV-based mmWave fronthaul for small cell networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 5560–5575, May 2023.

[11] M. T. Dabiri and M. Hasna, "3D uplink channel modeling of UAV-based mmWave fronthaul links for future small cell networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 1400–1413, Feb. 2023.

[12] Y. Hmamouche, M. Benjillali, S. Saoudi, H. Yanikomeroglu, and M. D. Renzo, "New trends in stochastic geometry for wireless networks: A tutorial and survey," *Proc. IEEE*, vol. 109, no. 7, pp. 1200–1252, Jul. 2021.

[13] M. R. Maleki, M. R. Mili, M. R. Javan, N. Mokari, and E. A. Jorswieck, "Multi-agent reinforcement learning trajectory design and two-stage resource management in comp UAV VLC networks," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7464–7476, Nov. 2022.

[14] Y. Zeng, X. Xu, S. Jin, and R. Zhang, "Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4205–4220, Jul. 2021.

[15] E. Fonseca, B. Galkin, R. Amer, L. A. DaSilva, and I. Dusparic, "Adaptive height optimisation for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Access*, vol. 11, pp. 5966–5980, 2023.

[16] L. Bellone, B. Galkin, E. Traversi, and E. Natalizio, "Deep reinforcement learning for combined coverage and resource allocation in UAV-aided RAN-slicing," 2022, *arXiv:2211.09713*.

[17] S. P. Gopi and M. Magarini, "Reinforcement learning aided UAV base station location optimization for rate maximization," *Electronics*, vol. 10, no. 23, p. 2953, 2021.

[18] S. A. Hoseini, J. Hassan, A. Bokani, and S. S. Kanhere, "Trajectory optimization of flying energy sources using Q-learning to recharge hotspot UAVs," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2020, pp. 683–688.

[19] Y.-J. Chen, W. Chen, and M.-L. Ku, "Trajectory design and link selection in UAV-assisted hybrid satellite-terrestrial network," *IEEE Wireless Commun. Lett.*, vol. 26, no. 7, pp. 1643–1647, Jul. 2022.

[20] Y.-J. Chen and D.-Y. Huang, "Joint trajectory design and BS association for cellular-connected UAV: An imitation-augmented deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2843–2858, Feb. 2022.

[21] Y.-J. Chen, K.-M. Liao, M.-L. Ku, F. P. Tso, and G.-Y. Chen, "Multi-agent reinforcement learning based 3D trajectory design in aerial-terrestrial wireless caching networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8201–8215, Aug. 2021.

[22] X. Zhang, H. Zhao, J. Wei, C. Yan, J. Xiong, and X. Liu, "Cooperative trajectory design of multiple UAV base stations with heterogeneous graph neural networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1495–1509, Mar. 2023.

[23] M. K. Shehzad, A. Ahmad, S. A. Hassan, and H. Jung, "Backhaul-aware intelligent positioning of UAVs and association of terrestrial base stations for fronthaul connectivity," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 4, pp. 2742–2755, Oct.–Dec. 2021.

[24] H. Lee, C. Eom, and C. Lee, "QoS-aware UAV-BS deployment optimization based on reinforcement learning," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, 2023, pp. 1–4.

[25] N. Hamden, A. Nasser, M. Y. Selim, and M. Elsabrouty, "Reinforcement learning based technique for interference management in UAV aided HetNets," in *Proc. 10th Int. Japan-Africa Conf. Electron., Commun., Comput. (JAC-ECC)*, 2022, pp. 81–84.

[26] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA: Wiley, 2016.

[27] M. T. Dabiri, H. Safi, S. Parsaeefard, and W. Saad, “Analytical channel models for millimeter wave UAV networks under hovering fluctuations,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2868–2883, Apr. 2020.

[28] M. T. Dabiri, S. M. S. Sadough, and M. A. Khalighi, “Channel modeling and parameter optimization for hovering UAV-based free-space optical links,” *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2104–2113, Sep. 2018.

[29] A. Kaadan, H. H. Refai, and P. G. LoPresti, “Multielement FSO transceivers alignment for inter-UAV communications,” *J. Lightw. Technol.*, vol. 32, no. 24, pp. 4785–4795, Dec. 15, 2014.

[30] T. Bai and R. W. Heath, “Coverage and rate analysis for millimeter-wave cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.

[31] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, “Modeling air-to-ground path loss for low altitude platforms in urban environments,” in *Proc. IEEE Global Commun. Conf.*, 2014, pp. 2898–2904.

[32] A. Al-Hourani, S. Kandeepan, and S. Lardner, “Optimal LAP altitude for maximum coverage,” *IEEE Wireless Commun. Let.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.

[33] R. Imran, M. Odeh, N. Zorba, and C. Verikoukis, “Quality of experience for spatial cognitive systems within multiple antenna scenarios,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4153–4161, Aug. 2013.

[34] N. Zorba and A. I. Perez-Neira, “Robust power allocation schemes for multibeam opportunistic transmission strategies under quality of service constraints,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 1025–1034, Aug. 2008.

[35] R. M. Nauss, “Solving the generalized assignment problem: An optimizing and heuristic approach,” *INFORMS J. Comput.*, vol. 15, no. 3, pp. 249–266, 2003.

[36] J. Achiam. “Twin delayed DDPG.” [Online]. Available: https://spinningup.openai.com/en/latest/algorithms/td3.html

[37] M. Rahmani et al., “Deep reinforcement learning-based sum rate fairness trade-off for cell-free mMIMO,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6039–6055, May 2023.

[38] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, “An overview of signal processing techniques for terahertz communications,” *Proc. IEEE*, vol. 109, no. 10, pp. 1628–1665, Oct. 2021.

[39] “Twin-delayed deep deterministic (TD3) policy gradient agents.” 2023. [Online]. Available: https://www.mathworks.com/help/reinforcement-learning/ug/td3-agents.html

[40] M. T. Dabiri and M. Hasna, “Pointing error modeling of mmWave to THz high-directional antenna arrays,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2435–2439, 2022.

[41] E. W. Weisstein. “Chi-squared distribution from MathWorld—A Wolfram Web resource.” [Online]. Available: https://mathworld.wolfram.com/Chi-SquaredDistribution.html