

# Meeting Stringent QoS Requirements in AAV-Assisted Networks: Resource Allocation and AAVs Positioning

MERIEM HAMMAMI<sup>1</sup>, CIRINE CHAIEB<sup>1</sup>, WESSAM AJIB<sup>1</sup> (Senior Member, IEEE), HALIMA ELBIAZE<sup>1</sup> (Senior Member, IEEE), AND ROCH GLITHO<sup>2</sup> (Life Senior Member, IEEE)

<sup>1</sup>Computer Sciences Department, University of Quebec at Montreal, Montreal, QC H3C 3P8, Canada

<sup>2</sup>Systems Engineering (CIISE), Concordia University, Montreal, QC H3G 1M8, Canada

CORRESPONDING AUTHOR: M. HAMMAMI (e-mail: meriemhammami934@gmail.com)

**ABSTRACT** Providing quick and reliable emergency communication in situations of natural disasters or unforeseen incidents may be crucial. In such situations, traditional communication infrastructure, such as ground-based wireless base stations, may become temporarily damaged or unavailable to support emergency teleoperations. Considered a promising solution, autonomous aerial vehicles (AAVs) can be deployed as flying base stations or relays to provide fast and reliable communication between physicians and remote robots in both uplink and downlink directions, while meeting strict transmission requirements. This paper addresses the joint optimization problem of AAV positioning and resource allocation in AAV-assisted wireless networks to minimize the number of deployed AAVs, all while satisfying stringent transmission quality demands. The formulated problem is a non-convex mixed-integer programming problem, which we prove to be  $\mathcal{NP}$ -hard. We first develop efficient greedy and metaheuristic genetic algorithms. Then, we propose an efficient centralized deep reinforcement learning solution based on the deep deterministic policy gradient (DDPG), where the agent learns optimal AAV positions and resource allocation. Simulation results demonstrate that the greedy solution closely matches the performance of both the genetic and deep reinforcement learning approaches, with a significant reduction in computational complexity. Furthermore, the results highlight the effectiveness of the deep reinforcement learning solution in minimizing the number of AAVs required to fully satisfy the transmission requirements of all users in both uplink and downlink directions.

**INDEX TERMS** AAV, mmWave communication, stringent applications, resource allocation, genetic algorithm, deep reinforcement learning.

## I. INTRODUCTION

**E**MERGENCY communication encompasses the methods and strategies used to facilitate rescue operations during natural disasters and manage communication needs during significant events such as holidays and conferences, when demand surges. The failure of ground communication infrastructure and associated systems, such as power facilities, often leads to widespread communication outages. This disruption impedes rescue efforts and hinders the rapid restoration of communication services. Additionally, communication outages make it extremely difficult to carry out operations such as emergency healthcare teleoperation. To enable such emergency communications over wireless

networks, various quality of service (QoS) requirements must be met, including a minimum data rate, minimal outage, and low latency. Therefore, ensuring efficient resource allocation solutions for both uplink (UL) and downlink (DL) information is critical [1].

Reestablishing communication systems is crucial for the effective exchange of information regarding the location, incidents, and severity of situations affecting both victims and rescuers. However, traditional emergency communication systems are based primarily on existing infrastructure, making it challenging to maintain communication services when this infrastructure is severely damaged. In such scenarios, conventional approaches for restoring communication

usually involve deploying emergency communication stations or repairing local base stations (BSs). Unfortunately, these methods tend to be time-consuming and logistically challenging to implement.

In emergency communication scenarios, satellite communications might be a potential solution to forward information between users and remote BSs [2], [3]. However, satellite communications are known for their high introductory costs, high latency, and insufficient throughput [4], which do not meet tele-operation requirements [5]. On the other hand, several wireless communication technologies have been developed to provide high coverage, low latency, and high data rates, such as Autonomous aerial vehicles (AAVs)-based communication. In the context of emergency healthcare tele-operation, where terrestrial network infrastructure can be wholly or partially absent, wireless emergency communication aided by AAVs is seen as a promising solution. It can break the isolation of disaster areas and allow physicians to perform first aid remotely thanks to AAVs' flexibility, mobility, low cost, and line-of-sight (LoS) links.

#### A. RELATED WORKS

Path planning is a critical aspect of designing AAV systems, as it involves determining AAVs' optimal positioning and trajectories. The trajectory of a AAV is influenced by various factors, such as flight duration, energy constraints, ground user demands, and the need to avoid collisions [6]. Optimizing AAV trajectories is a complex task that requires addressing multiple physical constraints and parameters. For instance, when determining the AAV trajectories for performance optimization, it is essential to take into account key factors such as channel variations caused by mobility, the dynamics of the AAV, energy consumption, and flight restrictions. Moreover, solving a continuous AAV trajectory optimization problem is analytically challenging. It involves identifying a potentially infinite number of optimization variables, specifically the AAV's locations. This process also requires understanding the interaction between mobility and different QoS metrics in wireless communication.

Several research works have focused on optimizing resource allocation and AAV locations in AAV-assisted wireless networks [7], [8], [9]. It exists different types of solutions to resolve this optimization problem as the mathematical solutions, heuristical solutions, metaheuristical solutions and machine learning-based solutions. In [10], the authors proposed a AAV-assisted Internet of Things (IoT) model for emergency communications to maximize the number of served IoT devices given AAVs' limited storage capacity. The problem optimized the power allocation and AAV trajectory while satisfying latency requirements. In [11], a cooperation scheme between AAVs was proposed to extend relay communication in a long-duration broadcast service context. The authors proposed a trajectory design and transmit power allocation strategy aimed at maximizing end-to-end throughput. The work in [12] studied AAV trajectory

optimization and bidirectional power allocation to maximize both UL and DL sum-rate. In [13], the authors studied a multi-user AAV-relaying system where each AAV acts as a relay for ground users. AAV placement and transmit power allocation were jointly optimized to maximize the achievable sum-rate of ground users. The authors in [14] proposed a AAV-aided full-duplex non-orthogonal multiple-access relay system. The joint optimization of UL and DL resource allocation was studied with the goal of minimizing the total power consumption. In [15], a hybrid-mode multiple access scheme was proposed for AAV-aided systems with heterogeneous traffic. The authors studied trajectory design and joint UL-DL bandwidth assignment to maximize the minimum average rate. The authors in [16] considered DL cell-free networks assisted by AAVs, where users can be associated with multiple AAVs. The joint optimization of AAVs positioning and resource allocation was studied to minimize the number of AAVs needed to satisfy users' requirements. The authors in [17] considered a multidrone cellular-connected IoD network, where drones transmit the collected data to the ground control station (GCS) using cellular RRBs. The envisioned system divides the set of transmitting drones into distinct drone pairs, where the paired drones simultaneously transmit over the same radio resource blocks (RRBs). These RRBs are shared with the terrestrial cellular network as well. The goal of this work was to maximize the uplink capacity of the IoD network while reducing interference over the shared RRBs between the IoD and cellular networks. Toward this goal, an optimization problem is presented to jointly perform drone pairing, transmit power allocation, and RRB scheduling among the drone pairs. The overall resource allocation is decomposed into three subproblems, namely, the drone transmit power allocation, drone-pairing and RRB scheduling and interference-aware RRB pricing. To obtain an efficient suboptimal solution, the authors proposed a solution of polynomial computational complexity for each subproblem.

Several works have investigated the utilization of meta-heuristic algorithms, such as genetic algorithms (GA), to tackle AAV trajectory design and resource allocation problems [18]. In [19], a AAV-enabled wireless power transfer system was considered, where the average received power among ground users was maximized by designing the optimal AAV trajectory. The authors in [20] optimized the AAV trajectory with the goal of minimizing transmit power while satisfying users' minimum data rate requirements. The work in [21] addressed post-disaster AAV-assisted communication, aiming to minimize the energy used to transfer data collected by BSs to the core network.

Deep reinforcement learning (DRL) techniques have recently been investigated and have proven their capability to efficiently tackle AAVs positioning and resource allocation problems in AAV-assisted wireless networks [22], [23]. Existing research works focused on different approaches (i.e., distributed, centralized, single agent, multi-agent) of the DRL techniques to solve this optimization problem.

In [24], a DRL-based solution was proposed to design the 3D AAV trajectory and frequency band allocation to fairly maximize user throughput. The proposed solution based on Deep Deterministic Policy Gradient (DDPG) allowed to enhance the energy efficiency by adjusting the flight speed and direction and to achieve fair communication service. The authors of [25] considered a system where the AAVs act as aerial BSs, and provide ubiquitous coverage. To maximize system utility across all served users, a problem combining joint user association, power allocation, and trajectory design has been addressed. The authors proposed a deep Q-learning (DQL) algorithm for trajectory design and resource allocation in a multi-AAV communications system, specifically focusing on downlink transmission networks. They also took into account the decentralized nature of multi-AAV systems and introduced a multi-agent DRL framework to develop a distributed algorithm, paving the way toward autonomous AAV communications systems. In [26], The authors aimed to ensure coverage continuity in high-dynamic AAV communication networks. To achieve this goal, they designed a resource allocation method based on deep reinforcement learning, which minimizes the rate variance of successive time slots. Their method adaptively adjusts neural network structures to meet coverage requirements by jointly allocating subchannels and power for ground users. To address the impact of channel state information (CSI) mismatch between the method's decisions and their implementation, the design of the reward function in the deep reinforcement learning solution incorporates temporal channel correlation. This approach helps reduce rate variance. The authors in [27] proposed an autonomous trajectory control method for multiple AAV BSs in wireless communications. Their multi-aerial BS trajectory control (MATC) scheme optimizes coverage and network throughput using a two-stage learning approach: a long short-term memory model estimates user link quality over time, while a centralized multi-agent deep reinforcement learning algorithm adjusts AAV trajectories for optimal communication performance. The authors in [28] considered a system of aerial backbone consisting of a connected group of AAVs to cover the integrality of the terrestrial vehicular environment while ensuring efficient data sharing and distribution. They proposed a strategy based on multi-agent Deep Q-Network (MA-DQN) to enhance the coverage of vehicles within the system. This approach involves simultaneously optimizing the flight paths and energy consumption of AAVs. The aim is to maximize the success rate of vehicles searching for data while also minimizing lookup latency. The authors in [29], a AAV data collector is dispatched to serve a set of IoT devices while being permanently followed and supported by an Unmanned Ground Vehicle (UGV) as a mobile charging station. Considering an unknown dynamic IoT environment, the problem of AoI minimization is reformulated as a Markov Decision Process (MDP). A MA-DQN method was employed to minimize the average AoI of IoT devices, timely recharge the AAV using UGV, and optimize the

UGV trajectory on the ground by considering the terrestrial obstacles and the IoT devices' movements. The authors in [30] studied a cellular-connected Internet of Drones (IoD) network with full-duplex cellular base stations (CBSs) using orthogonal resource blocks for aerial communication. CBSs connect to drone clusters via uplink NOMA and transmit artificial noise to counter eavesdropping to enhance security. They proposed a resource allocation scheme to manage interference and enhance physical layer security against multi-band eavesdropping drones with the goal of maximizing the worst-case average sum-secrecy rate of the network. The optimization problem, addressing drone clustering, transmit power, and jamming power allocation, is solved using a multi-agent reinforcement learning framework for clustering and successive convex approximation for power allocation. In [31], the authors considered AAV-assisted emergency communication, where each AAV acts as a mobile relay to forward data from a macro BS to users. They jointly studied power allocation, AAV service zone selection, and user scheduling. A DRL-based solution was proposed to maximize sum spectrum efficiency. The authors in [32] proposed a collaborative multi-agent DRL approach to maximize the utility of an entire multi-AAV network. Each AAV was modeled as an independent agent capable of choosing its deployment position, transmission power, and resource selection. The work in [33] studied the joint design of AAV positions, transmit beamforming, and AAV-user association in AAV-assisted DL cellular networks to maximize the achievable sum rate. The study proposed combining two techniques to solve the joint optimization problem: first, a DRL-based solution to address the issue of CSI unavailability for AAV position design, and second, a difference of convex algorithm to solve AAV transmit beamforming and AAV-user association efficiently. The authors in [34] considered a DL cellular network where multiple AAVs serve as aerial BSs for ground users and coordinate decision-making, including resource allocation and trajectory design, in a decentralized manner. To optimize overall fair throughput, they proposed a distributed DRL framework based on parameterized deep Q-learning to handle mixed action spaces (i.e., discrete and continuous actions). A comparative analysis of the proposed work with recent related studies is summarized in Table 1.

We note that most of the research that optimizes AAV resource allocation and trajectory (i.e., positioning) in a multi-AAV network proposes distributed multi-agent or DQN-based reinforcement learning solutions. Taking into account the dynamics of our system, we believe that a centralized DRL solution based on a single agent is best suited to guarantee convergence. Deep Q-Network (DQN) is a reinforcement learning algorithm that is well-suited for problems where both the state and action spaces are discrete. However, in our problem, the action space is continuous, which presents a significant challenge for DQN. Although proximal policy optimization (PPO) is simpler and more stable than DDPG, it is less sample-efficient. One of the key

TABLE 1. A comparative analysis of the proposed work with existing studies.

Ref	Objective function				Link		UAV position		QoS	
	rate	IoT connections	energy	number of UAVs	UP	DL	discrete	continuous	rate	outage
[10]		✓			✓	✓		✓	✓	
[11], [13], [31]	✓				✓	✓	✓		✓	
[12]	✓				✓	✓		✓	✓	
[14], [28]			✓		✓	✓	✓		✓	
[15]	✓				✓	✓		✓	✓	
[35]				✓		✓		✓	✓	
[19], [20]		✓				✓		✓	✓	
[21]		✓			✓			✓	✓	
[24], [33], [34]	✓					✓		✓	✓	
[29]	✓					✓	✓		✓	
[25], [27], [26]	✓					✓	✓		✓	
our paper				✓	✓	✓	✓	✓	✓	✓

differences between the two is that DDPG is off-policy, while PPO is on-policy. DDPG focuses on selecting the overall best policy to maximize rewards by consistently learning from experiences and exploration, whereas PPO relies on the current training policy, regardless of how it compares to previous policies. For these reasons, we chose DDPG over PPO and DQN as the most suitable approach for this work.

**B. CONTRIBUTIONS**

The joint optimization problem of AAVs positioning and resource allocation has been tackled in several works, either in the UL or DL direction [36]. To the best of our knowledge, minimizing the number of AAVs needed to meet all stringent emergency communications requirements in both UL and DL directions has not yet been addressed. Most existing works have focused on maximizing end-to-end throughput or enhancing the achievable sum-rate for ground users. In contrast, some studies targeting energy efficiency have optimized total power consumption in AAV-aided systems.

Our paper addresses this gap by addressing the AAV positioning and resource allocation problem while ensuring minimum rate and outage probability thresholds. As in [37], our model considers outage probability thresholds, since it is not always possible to ensure stringent SINR (signal-to-noise ratio) or delay thresholds due to potential deep fading situations in a wireless environment. We studied AAVs positioning and resource allocation problem (UPRAP) in a AAV-assisted wireless network for emergency tele-operation communications. Our main contributions are summarized as follows.

- We mathematically formulate the joint UPRAP, which includes AAV positioning, power allocation, and user association, and prove its  $\mathcal{NP}$ -hardness. The QoS requirements for each user are defined by the need to meet minimum rate and outage probability thresholds for both UL and DL links.

- To solve the problem efficiently, we propose a greedy-based polynomial algorithm to select discrete positions for the deployed AAVs, associate users with the AAVs, and allocate transmit power levels. Since GA is known to converge to an optimal or near-optimal solution after a certain number of iterations, we also propose a metaheuristic genetic algorithm.
- To take advantage of the continuous action space of deep deterministic policy gradient (DDPG), we develop a centralized DRL algorithm based on DDPG approach. UPRAP is modeled as a Markov Decision Process (MDP), where the actions involve selecting positions.
- We evaluate the performance of the proposed algorithms through simulations by exploring: (i) the feasibility region of the problem, defined as the range of user requirements for which all users can be satisfied; (ii) the impact of system parameters (e.g., number of channels, required rate and outage probability); and (iii) the impact of imperfect knowledge of CSI. The results demonstrate the effectiveness of the DRL-based solution and highlight the effect of AAV position discretization on the number of AAVs required for deployment.

**C. ORGANIZATION**

The rest of the paper is organized as follows. Section II summarizes the system model. Section III formulates the optimization problem and studies its complexity. The proposed greedy and genetic-based solutions are described, respectively, in Sections IV and V. Section VI details the DRL-based solution and analyses its complexity. Section VII presents and discusses the simulation results and finally Section VIII concludes the paper.

**II. SYSTEM MODEL**

Consider a AAV-assisted wireless network where a set  $\mathcal{N}$  of  $N$  AAVs can fly over an area covered by a macro BS (MBS) located at its center, serving a set  $\mathcal{U}$  of  $U$  users, representing

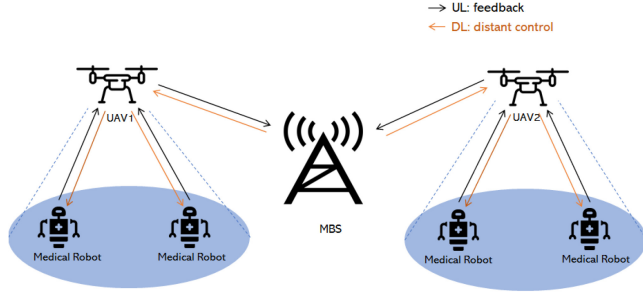


FIGURE 1. System model.

remotely controlled medical robots. As shown in Fig. 1, the AAVs act as relays between the MBS and the ground users. The MBS, as a central unit, is responsible for preventing AAV collisions and ensuring a reasonable distance between AAVs. The available bandwidth for UL communications is assumed to be equal to that for DL communications. Let  $B_{fh}$  and  $B_{bh}$  represent the available bandwidth in front-haul links (i.e., between users and AAVs) and back-haul links (i.e., between AAVs and the MBS), respectively. Without loss of generality, we assume identical channels and that the number of UL channels,  $C_{ul}$ , is equal to the number of DL channels,  $C_{dl}$ , so  $C = C_{ul} = C_{dl}$ . Therefore,  $B_{fh}$  is divided into  $C$  channels for each direction.

The deployed AAVs are assumed to fly at a fixed altitude  $H_{uav}$ . Let  $q_n = (x_n, y_n, H_{uav})$ ,  $q_u = (x_u, y_u, 0)$  and  $q_{bs} = (x_{BS}, y_{BS}, H_{bs})$  denote respectively the positions of AAV  $n$ , user  $u$  and the MBS, where  $H_{bs}$  is the BS altitude. Let  $\|\cdot\|$  denotes the two-norm, the distance between AAV  $n$  and user  $u$  and the distance between AAV  $n$  and the MBS are denoted by  $d_{n,u} = d_{u,n} = \|q_n - q_u\|$  and  $d_{BS,n} = d_{n,BS} = \|q_{bs} - q_n\|$ , respectively. For clarity, we summarize the used notations in Table 2.

### A. CHANNEL MODEL

Similarly to [38], [39], we consider a *Nakagami- $m$*  fading environment. Hence, the small-scale channel coefficient, denoted by  $h_\diamond$ , with  $\diamond \in \{(n, u), (u, n), (n, BS), (BS, n)\}$ , follows *Nakagami- $m$*  distribution. The mmWave small scale channel gain  $|h_\diamond|^2 \sim \Gamma(m_\star, 1/m_\star)$  is an independent and identically distributed Gamma random variable where  $\star \in \{\text{LoS}, \text{NLoS}\}$  and  $m_\star$  are the *Nakagami* parameters for LoS and NLoS links. The mean of  $|h_\diamond|^2$  is  $E(|h_\diamond|^2) = \mu = k\theta$  and depends on its parameters  $k$  and  $\theta$ . In this case,  $E(|h_\diamond|^2) = \mu = k\theta = m_\star \cdot 1/m_\star = 1$ . We assume that users, AAVs, and MBS are equipped with beamforming antenna arrays, where a widely used sector antenna model is considered. The beamforming technique enhances the signal strength at the receiver and reduces interferences with other equipment as it dynamically adjusts the phase and amplitude of the signal at each antenna to steer the beam in a desired direction. Let  $G_m$  and  $G_s$  denote the main and the side lobes, and the

TABLE 2. List of symbols and definitions.

Symbols	Definitions
$\mathcal{N}, N$	Set and number of UAVs
$\mathcal{U}, U$	Set and number of users
$\mathcal{C}_{ul}, C_{ul}$	Set and number of available UL channels
$\mathcal{C}_{dl}, C_{dl}$	Set and number of available DL channels
$q_n$	Position of UAV $n$
$q_u$	Position of user $u$
$q_{bs}$	Position of the MBS
$d_{n,BS} = d_{BS,n}$	Distance between the MBS and UAV $n$
$d_{n,u} = d_{u,n}$	Distance between the MBS and UAV $n$
$h_{BS,n}$	Small-scale DL channel coefficient of MBS-UAV $n$ link
$h_{n,BS}$	Small-scale UL channel coefficient of MBS-UAV $n$ link
$h_{n,u}$	Small-scale UL channel coefficient of user $u$ -UAV $n$ link
$h_{n,u}$	Small-scale DL channel coefficient of user $u$ -UAV $n$ link
$g_{n,BS}$	DL channel gain between the MBS and UAV $n$
$g_{n,BS}$	UL channel gain between the MBS and UAV $n$
$g_{u,n}$	UL channel gain between user $u$ and UAV $n$
$g_{u,n}$	DL channel gain between user $u$ and UAV $n$
$e_{BS,n}$	DL channel estimation error between the MBS and UAV $n$
$e_{n,BS}$	UL channel estimation error between the MBS and UAV $n$
$e_{u,n}$	UL channel estimation error between user $u$ and UAV $n$
$e_{n,u}$	DL channel estimation error between user $u$ and UAV $n$
$\sigma_e^2$	Variance of the channel estimation error
$I_{u,n,c_{ul}}$	Interference received at UAV $n$ associated with user $u$ over $c_{ul}$
$I_{n,u,c_{dl}}$	Interference received at user $u$ associated with UAV $n$ over $c_{dl}$
$\mathbf{b}$	The UAVs activation vector
$\mathbf{Z}$	The association matrix
$p_{u,n}$	Transmit power level from user $u$ to UAV $n$
$p_{n,u}$	Transmit power level from UAV $n$ to user $u$
$p_{BS,n}$	Transmit power level from MBS to UAV $n$
$p_{n,BS}$	Transmit power level from UAV $n$ to MBS
$R_{n,u,c}, O_{n,u}$	DL rate between UAV $n$ and user $u$ and its outage probability
$R_{u,n,c}, O_{u,n}$	UL rate between UAV $n$ and user $u$ and its outage probability
$R_{BS,n}, O_{BS,n}$	DL rate between UAV $n$ and MBS and its outage probability
$R_{n,BS}, O_{n,BS}$	UL rate between UAV $n$ and MBS and its outage probability

antenna gain  $G_\diamond$  is modeled as [40]:

$$G_\diamond = \begin{cases} G_m & \text{with probability } \theta/2\pi \\ G_s & \text{with probability } (2\pi - \theta)/2\pi, \end{cases} \quad (1)$$

where  $\theta$  is the width of the transmitting antenna.

Finally, the mmWave channel gain  $g_\diamond$  is given by  $g_\diamond = G_\diamond f^{-1}(d_\diamond) |h_\diamond|^2$  where  $f$  is the path-loss (PL) function expressed as follows:

$$10 \log_{10}(f(d_\diamond)) = P_{LoS} PL^{LoS}(d_\diamond) + (1 - P_{LoS}) PL^{NLoS}(d_\diamond), \quad (2)$$

where  $P_{LoS}$  is the LoS probability [41] and  $d_\diamond$  is the distance. The PL function is given by [42]:

$$PL^*(d_\diamond) = \alpha^* + 10\beta^* \log_{10}(d_\diamond) + \xi \text{ [dB]}, \quad \xi \sim \mathcal{N}(0, \sigma_\xi^{*2}), \quad (3)$$

where  $\alpha^*$ ,  $\beta^*$ , and  $\sigma_\xi^*$  are the least square fits of floating intercept, the slope over the measured distances, and the lognormal shadowing variance, respectively.

### B. CHANNEL MODEL WITH IMPERFECT CSI

In order to run centralized resource allocation algorithms, we assume that the MBS has a perfect knowledge of CSI. However, due to channel fluctuations, limited feedback, channel estimation, and quantization errors, a perfect knowledge of CSI can be challenging to obtain. We assume that the CSI is imperfectly known and the estimated channel gain at the destination is given by [43]:

$$\tilde{h}_\diamond = h_\diamond + e_\diamond, \quad (4)$$

where  $e_\diamond$  is the channel estimation error which follows zero-mean complex Gaussian distribution with variance  $\sigma_e^2$  (i.e.,  $e_\diamond \sim \mathcal{CN}(0, \sigma_e^2)$ ). Hence, the mmWave channel gain with imperfect CSI is  $\tilde{g}_\diamond = G_\diamond f^{-1}(d_\diamond) |\tilde{h}_\diamond|^2$ .

### C. TRANSMISSION MODEL

To model the optimization problem, we introduce the following binary variables:

$$b_n = \begin{cases} 1 & \text{if AAV } n \text{ is deployed} \\ 0 & \text{otherwise} \end{cases}, \text{ and} \quad (5)$$

$$z_{u,n,c_{ul},c_{dl}} = \begin{cases} 1 & \text{if } u \text{ is associated with } n \text{ over } c_{ul}, c_{dl} \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Since the mmWave bandwidth is shared by all AAVs, the transmitted signal from user  $u$  to AAV  $n$  over channel  $c_{ul}$  would be subject to interference which can be expressed as:

$$I_{u,n,c_{ul}} = \sum_{u' \in \mathcal{U}'} \sum_{n' \in \mathcal{N}'} \sum_{c_{dl} \in \mathcal{C}_{dl}} P_{u',n'} g_{u',n} z_{u',n',c_{ul},c_{dl}} \quad (7)$$

We assume that the associated user is in the main lobe of its serving AAV. We suppose that UL and DL channels are perfectly orthogonal. The received interference at user  $u$  that is associated with AAV  $n$  and using channel  $c_{dl}$  can be given as follows:

$$I_{n,u,c_{dl}} = \sum_{u' \in \mathcal{U}'} \sum_{n' \in \mathcal{N}'} \sum_{c_{ul} \in \mathcal{C}_{ul}} P_{n',u'} g_{n',u} z_{u',n',c_{ul},c_{dl}}, \quad (8)$$

where  $\mathcal{U}' = \mathcal{U} \setminus \{u\}$  and  $\mathcal{N}' = \mathcal{N} \setminus \{n\}$ .

The expected values, denoted by  $\mathbb{E}$ , of the AAV-user interference in UL and DL links can be respectively given by:

$$\mathbb{E}(I_{u,n,c_{ul}}) = \sum_{u' \in \mathcal{U}'} \sum_{n' \in \mathcal{N}'} \sum_{c_{dl} \in \mathcal{C}_{dl}} z_{u',n',c_{ul},c_{dl}} P_{u',n'} G_{u',n} f^{-1}(d_{u',n}), \quad (9)$$

and

$$\mathbb{E}(I_{n,u,c_{dl}}) = \sum_{u' \in \mathcal{U}'} \sum_{n' \in \mathcal{N}'} \sum_{c_{ul} \in \mathcal{C}_{ul}} z_{u',n',c_{ul},c_{dl}} P_{n',u'} G_{n',u} f^{-1}(d_{u',n'}). \quad (10)$$

Finally, the achievable UL data rate from user  $u$  to AAV  $n$  over channel  $c_{ul}$  is calculated as follows:

$$R_{u,n,c_{ul}} = W_{fh} \log_2 \left( 1 + \frac{P_{u,n} g_{u,n}}{I_{u,n,c_{ul}} + N_0} \right), \quad (11)$$

where  $W_{fh} = B_{fh}/C$  is the front-haul channel bandwidth,  $N_0 = \sigma^2 W_{fh}$  is the additive white Gaussian noise (AWGN) power, and  $\sigma^2$  is the AWGN power spectral density,  $p_{u,n}$  is the transmit power. The achievable DL data rate from AAV  $n$  to user  $u$  over channel  $c_{dl}$ , denoted by  $R_{n,u,c_{dl}}$ , can be calculated by substituting respectively  $u, n$  and  $c_{ul}$  in (11) by  $n, u$  and  $c_{dl}$ .

The back-haul and front-haul channels are assumed perfectly orthogonal. The achievable UL data rate between AAV  $n$  and the MBS can be given by:

$$R_{n,BS} = W_{bh} \log_2 \left( 1 + \frac{p_{n,BS} |h_{n,BS}|^2 f^{-1}(d_{n,BS}) G_m}{N'_0} \right), \quad (12)$$

where  $W_{bh} = B_{bh}/\sum_{n \in \mathcal{N}} b_n$  is the back-haul bandwidth,  $p_{n,BS}$  is the transmit power and  $N'_0 = \sigma^2 W_{bh}$ . The achievable DL data rate from MBS to AAV  $n$  denoted by  $R_{BS,n}$ , can be calculated by substituting  $n, BS$  in (12) by  $BS, n$ .

A user is, by definition, considered in outage if its transmission rate is lower than a given threshold rate. Similarly to [44], the UL outage probability of user  $u$  and its associated AAV  $n$ , denoted by  $O_{u,n}$ , can be given by:

$$\begin{aligned} O_{u,n} &= \mathbb{P}(R_{u,n,c_{ul}} \leq \zeta_{fh}) \\ &= \mathbb{P} \left( \frac{p_{u,n} |h_{u,n}|^2 f^{-1}(d_{u,n}) G_m}{I_{u,n,c_{ul}} + N_0} \leq \gamma_{th,u,n} \right) \\ &\approx P_{LoS} F \left( \frac{\gamma_{th,u,n} (\mathbb{E}(I_{u,n,c_{ul}}) + N_0)}{p_{u,n} G_m PL^{LoS}(d_{u,n})} \right) \\ &\quad + P_{NLoS} F \left( \frac{\gamma_{th,u,n} (\mathbb{E}(I_{u,n,c_{ul}}) + N_0)}{p_{u,n} G_m PL^{NLoS}(d_{u,n})} \right) \\ &\approx P_{LoS} \frac{1}{\Gamma(m_{LoS})} \\ &\quad \gamma \left( m_{LoS}, \frac{m_{LoS} \gamma_{th,u,n} (\mathbb{E}(I_{u,n,c_{ul}}) + N_0)}{p_{u,n} G_m PL^{LoS}(d_{u,n})} \right) \\ &\quad + P_{NLoS} \frac{1}{\Gamma(m_{NLoS})} \\ &\quad \gamma \left( m_{NLoS}, \frac{m_{NLoS} \gamma_{th,u,n} (\mathbb{E}(I_{u,n,c_{ul}}) + N_0)}{p_{u,n} G_m PL^{NLoS}(d_{u,n})} \right), \quad (13) \end{aligned}$$

where  $\zeta_{fh}$  is the threshold rate required for both UL and DL transmissions in front-haul links,  $\gamma_{th,u,n} = 2^{\zeta_{fh}/W_{fh}} - 1$  is the SINR threshold,  $F$  is the cumulative distribution function (CDF),  $\Gamma$  is the gamma function, and finally  $\gamma$  is the lower incomplete gamma function. Similarly to (13), the DL outage probability between AAV  $n$  and user  $u$ , denoted by  $O_{n,u}$ , can be expressed as follows:

$$\begin{aligned} O_{n,u} &\approx P_{LoS} \frac{1}{\Gamma(m_{LoS})} \\ &\quad \gamma \left( m_{LoS}, \frac{m_{LoS} \gamma_{th,n,u} (\mathbb{E}(I_{n,u,c_{dl}}) + N_0)}{p_{n,u} G_m PL^{LoS}(d_{u,n})} \right) \\ &\quad + P_{NLoS} \frac{1}{\Gamma(m_{NLoS})} \\ &\quad \gamma \left( m_{NLoS}, \frac{m_{NLoS} \gamma_{th,n,u} (\mathbb{E}(I_{n,u,c_{dl}}) + N_0)}{p_{n,u} G_m PL^{NLoS}(d_{u,n})} \right), \quad (14) \end{aligned}$$

where  $\gamma_{th,n,u} = 2^{\zeta_{fh}/W_{fh}} - 1$  is the SINR threshold.

The outage probability between AAV  $n$  and the MBS, denoted by  $O_{n,BS}$ , can be expressed as follows:

$$\begin{aligned}
O_{n,BS} &= \mathbb{P}(R_{n,BS} \leq \zeta_{bh}) \\
&= \mathbb{P}\left(\frac{p_{n,BS}|h_{n,BS}|^2 f^{-1}(d_{n,BS})G_m}{N'_0} \leq \gamma_{th,n,BS}\right) \\
&= P_{LoS}F\left(\frac{\gamma_{th,n,BS}N'_0}{p_{n,BS}G_m PL^{LoS}(d_{n,BS})}\right) \\
&\quad + P_{NLoS}F\left(\frac{\gamma_{th,n,BS}N'_0}{p_{n,BS}G_m PL^{NLoS}(d_{n,BS})}\right) \\
&= P_{LoS} \frac{1}{\Gamma(m_{LoS})} \\
&\quad \gamma\left(m_{LoS}, \frac{m_{LoS}\gamma_{th,n,BS}N'_0}{p_{n,BS}G_m PL^{LoS}(d_{n,BS})}\right) \\
&\quad + P_{NLoS} \frac{1}{\Gamma(m_{NLoS})} \\
&\quad \gamma\left(m_{NLoS}, \frac{m_{NLoS}\gamma_{th,n,BS}N'_0}{p_{n,BS}G_m PL^{NLoS}(d_{n,BS})}\right), \quad (15)
\end{aligned}$$

where  $\zeta_{bh}$  is the required threshold rate for both UL and DL back-haul links,  $\gamma_{th,n,BS} = 2^{\zeta_{bh}/W_{bh}} - 1$  is the UL signal to noise ratio (SNR) threshold between AAV  $n$  and the MBS. Similarly to (15), the DL outage probability between AAV  $n$  and the MBS, denoted by  $O_{BS,n}$ , is given by:

$$\begin{aligned}
O_{BS,n} &= P_{LoS} \frac{1}{\Gamma(m_{LoS})} \\
&\quad \gamma\left(m_{LoS}, \frac{m_{LoS}\gamma_{th,BS,n}N'_0}{p_{BS,n}G_m PL^{LoS}(d_{n,BS})}\right) \\
&\quad + P_{NLoS} \frac{1}{\Gamma(m_{NLoS})} \\
&\quad \gamma\left(m_{NLoS}, \frac{m_{NLoS}\gamma_{th,BS,n}N'_0}{p_{BS,n}G_m PL^{NLoS}(d_{n,BS})}\right), \quad (16)
\end{aligned}$$

where  $\gamma_{th,BS,n} = 2^{\zeta_{bh}/W_{bh}} - 1$  is the SNR threshold.

### III. JOINT RESOURCE ALLOCATION AND AAVS POSITIONING PROBLEM

This section formulates the UPRAP as a mixed-integer non-linear programming problem (MINLP) and proves its  $\mathcal{NP}$ -hardness.

#### A. PROBLEM FORMULATION

The objective of UPRAP is to minimize the number of deployed AAVs while ensuring emergency operations under strict constraints on outage probabilities and transmission power budgets in each time slot. Let  $\mathbf{b} = [b_n]$  be the binary  $N$ -size vector to indicate the deployed AAVs and  $\mathcal{P} = \{\mathbf{P}_{user}, \mathbf{P}_{AAV}^{dl}, \mathbf{P}_{AAV}^{ul}, \mathbf{P}_{BS}\}$  be the set of power levels of the  $U \times N$  matrix  $\mathbf{P}_{user} = [p_{u,n}]$ , the  $N \times U$  matrix  $\mathbf{P}_{AAV}^{dl} = [p_{n,u}]$ , and the  $N$ -size vectors  $\mathbf{P}_{AAV}^{ul} = [p_{n,BS}]$  and  $\mathbf{P}_{BS} = [p_{BS,n}]$ . We also denote the  $U \times N \times C_{ul} \times C_{dl}$  association matrix by  $\mathbf{Z} = [z_{u,n,c_{ul},c_{dl}}]$  and the AAVs'

location  $N$ -size vector by  $\mathbf{Q} = [q_n]$ . The optimization problem is mathematically formulated as follows:

$$\text{minimize}_{\mathbf{Z}, \mathbf{Q}, \mathbf{b}, \mathcal{P}} \sum_{n \in \mathcal{N}} b_n \quad (\text{P1a})$$

$$\text{s.t. } b_n \in \{0, 1\}, \forall n \in \mathcal{N}, \quad (\text{P1b})$$

$$z_{u,n,c_{ul},c_{dl}} \in \{0, 1\}, \quad (\text{P1c})$$

$$\forall n \in \mathcal{N}, u \in \mathcal{U}, c_{ul} \in \mathcal{C}_{ul}, c_{dl} \in \mathcal{C}_{dl}, \quad (\text{P1c})$$

$$\sum_{n \in \mathcal{N}} p_{BS,n} b_n \leq p_{BS}^{max}, \quad (\text{P1d})$$

$$b_n \sum_{u \in \mathcal{U}} \sum_{c_{dl} \in \mathcal{C}_{dl}} \sum_{c_{ul} \in \mathcal{C}_{ul}} p_{n,u} z_{u,n,c_{ul},c_{dl}} + p_{n,BS} \leq p_n^{max}, \quad (\text{P1e})$$

$$p_{u,n} z_{u,n,c_{ul},c_{dl}} \leq p_u^{max}, \quad (\text{P1f})$$

$$n \in \mathcal{N}, u \in \mathcal{U}, c_{ul} \in \mathcal{C}_{ul}, c_{dl} \in \mathcal{C}_{dl}, \quad (\text{P1f})$$

$$\sum_{n \in \mathcal{N}} \sum_{c_{ul} \in \mathcal{C}_{ul}} \sum_{c_{dl} \in \mathcal{C}_{dl}} z_{u,n,c_{ul},c_{dl}} = 1, \forall u \in \mathcal{U}, \quad (\text{P1g})$$

$$z_{u,n,c_{ul},c_{dl}} O_{n,u} \leq \hat{O}_{fh} \text{ and } z_{u,n,c_{ul},c_{dl}} O_{u,n} \leq \hat{O}_{bh}, \quad (\text{P1h})$$

$$\forall n \in \mathcal{N}, u \in \mathcal{U}, c_{ul} \in \mathcal{C}_{ul}, c_{dl} \in \mathcal{C}_{dl}, \quad (\text{P1h})$$

$$b_n O_{n,BS} \leq \hat{O}_{bh} \text{ and } b_n O_{BS,n} \leq \hat{O}_{bh}, \forall n \in \mathcal{N}, \quad (\text{P1i})$$

$$\sum_{u \in \mathcal{U}} \sum_{c_{dl} \in \mathcal{C}_{dl}} z_{u,n,c_{ul},c_{dl}} \leq 1, \forall n \in \mathcal{N}, c_{ul} \in \mathcal{C}_{ul}, \quad (\text{P1j})$$

$$\sum_{u \in \mathcal{U}} \sum_{c_{ul} \in \mathcal{C}_{ul}} z_{u,n,c_{ul},c_{dl}} \leq 1, \forall n \in \mathcal{N}, c_{dl} \in \mathcal{C}_{dl}, \quad (\text{P1k})$$

$$\sum_{u \in \mathcal{U}} \sum_{c_{ul} \in \mathcal{C}_{ul}} \sum_{c_{dl} \in \mathcal{C}_{dl}} z_{u,n,c_{ul},c_{dl}} \leq b_n, \forall n \in \mathcal{N}, \quad (\text{P1l})$$

$$b_n - z_{u,n,c_{ul},c_{dl}} \geq 0, \quad (\text{P1m})$$

$$\forall n \in \mathcal{N}, u \in \mathcal{U}, c_{ul} \in \mathcal{C}_{ul}, c_{dl} \in \mathcal{C}_{dl}, \quad (\text{P1m})$$

$$q_n b_n \neq q_{n'} b_{n'}, \quad \forall n, n' \in \mathcal{N}, n \neq n'. \quad (\text{P1n})$$

Constraints (P1d) and (P1e) restrict the total transmission power of the MBS and AAVs to not exceed the maximum available power budgets  $p_{BS}^{max}$  and  $p_u^{max}$ , respectively. Constraints (P1f) limit the transmit power of the users.

Constraints (P1g) ensure that user  $u$  can be at most associated with one AAV over a single channel in the UL and a single channel in the DL. Reliable services and low-outage probabilities are guaranteed by constraints (P1h) and (P1i) where  $\hat{O}_{bh}$  and  $\hat{O}_{fh}$  are the outage probability thresholds of back-haul and front-haul links, respectively. Constraints (P1j) and (P1k) ensure that each UL and DL channel is used by only one user. Constraints (P1l) and (P1m) express the binary relationship between the AAVs' deployment and the users association. Constraints (P1n) ensure that two AAVs can not be placed in the same position. (P1) is a non-convex and mixed-integer and non-linear programming problem due to the logarithmic function and the constraints that involve the decision matrix  $\mathbf{Z}$  and decision vector  $\mathbf{b}$ . Consequently, (P1) is nontractable.

## B. PROBLEM COMPLEXITY

In the following, the  $\mathcal{NP}$ -hardness of (P1) is investigated by reducing the facility location problem (FLP), which is known to be an  $\mathcal{NP}$ -hard problem, to a special case of UPRAP. The following lemma proves this result [45].

*Lemma 1:* UPRAP formulated in (P1) is  $\mathcal{NP}$ -hard.

*Proof:* We prove that UPRAP is  $\mathcal{NP}$ -hard by restriction, that is, we show that UPRAP contains FLP as a special case. In FLP, we are given a set of  $\mathcal{J} = \{j, \dots, J\}$  of cities, a set  $\mathcal{F} = \{i, \dots, F\}$  of potential locations for facilities. The cost of opening a facility at location  $i \in \mathcal{F}$ , if it is used to serve a number of cities, is  $f_i$ . The cost of connecting a city  $j \in \mathcal{J}$  to facility  $i$  is  $c_{ji}$ . Further, the FLP problem is usually formulated as a binary integer program with binary variables.  $x_{ji}$  indicates whether city  $j$  is connected to facility  $i$  and  $y_i$  indicates whether facility  $i$  is opened or not. The main purpose of FLP is to decide where to place a number of facilities, in order to cover the need of service for a number of cities, in the most efficient way by minimizing the costs of the placement of the different facilities.

Given an instance of FLP, an instance of UPRAP can be constructed as follows:  $F = N$ ,  $J = U$ ,  $c_{ji} = 0$ ,  $f_i = 1$  if AAV  $i$  is deployed,  $x_{ji} = z_{j,i,c_{ul},c_{dl}}$  and  $y_i = b_i \forall i \in \mathcal{N}, j \in \mathcal{U}$ . We assume that there are no power restrictions (constraints (P1d), (P1e) and (P1f)), thus all QoS requirements can be met. We also assume that each user can be associated to multiple AAVs (constraints (P1g)), each channel can be used by multiple users (constraints (P1j) and (P1k)), and finally, the AAVs can take the same position but at different altitudes (constraints (P1n)). These assumptions imply that the mentioned constraints are not applicable. Under this restriction, UPRAP is equivalent to minimizing the number of deployed AAVs to serve all the users subject to QoS constraints. By analogy, if AAVs represent the facilities, the users represent the cities, the cost of deploying AAV  $i$  is  $f_i = 1$ , and the cost of associating user  $j$  to AAV  $i$  is  $c_{ji} = 0$ . UPRAP becomes equivalent to FLP. Since FLP is  $\mathcal{NP}$ -hard, then so is UPRAP. This proves the lemma. ■

It is to be reminded that our problem should be solved each time slot since the channel states change every time slot. It is challenging to optimize both the association and the AAV positions in each time slot and a more practical solution could be to consider two time-scale optimization (e.g., optimizing the AAV positions on a longer timescale). Anyhow, for simplicity purposes, this paper considers one time-scale optimization that could be seen as a perfect case. Solving (P1) is not mathematically straightforward and is computationally expensive since all possible combinations of user-AAV association, power allocation, and AAVs positions solutions have to be carefully evaluated. Hence, efficient algorithmic solutions are discussed in the next three sections.

## IV. GREEDY SOLUTION

This section details the proposed greedy algorithm. We assume that the MBS knows the CSI between all users and

### Algorithm 1 UDPA Algorithm

---

**Initialize:**  $\mathbf{P}_{u,n} \leftarrow \mathbf{0}$ ,  $\mathbf{P}_{n,u} \leftarrow \mathbf{0}$ ,  $\mathbf{p}_{n,BS} \leftarrow \mathbf{0}$ ,  $\mathbf{p}_{BS,n} \leftarrow \mathbf{0}$ ,  
 $\mathbf{Z} \leftarrow \mathbf{0}$ ,  $\mathbf{b} \leftarrow \mathbf{0}$ ,  $p_{BS} \leftarrow p_{BS}^{max}$ ,  $p_n \leftarrow p_n^{max}$ ,  $\mathcal{U}_{nas} \leftarrow \mathcal{U}$ .

- 1:  $n \leftarrow 0$
- 2: **while**  $\mathcal{U}_{nas}$  is not empty **do**
- 3:    $n \leftarrow n + 1$
- 4:   **for** each position  $q$  in the set of positions  $\mathcal{Q}$  **do**
- 5:     Allocate minimum  $p_{n,BS}$  and  $p_{BS,n}$  to satisfy (P1i).
- 6:     **if** (P1d) and (P1i) **then**
- 7:       Update  $p_{BS}$  and  $p_n$ .
- 8:       Sort  $\mathcal{U}_{nas}$  according to decreasing channel gains.
- 9:       **for**  $u \in \mathcal{U}_{nas}$  **do**
- 10:         **if**  $c_{ul} \in \mathcal{C}_{ul}$  and  $c_{dl} \in \mathcal{C}_{dl}$  are available **then**
- 11:         Allocate min  $p_{u,n}$  and  $p_{n,u}$  to satisfy (P1h).
- 12:         **if** (P1e) and (P1h) **then**
- 13:         Associate user  $u$  to AAV  $n$ .
- 14:          $\mathcal{U}_{nas} \leftarrow \mathcal{U}_{nas} \setminus \{u\}$ .
- 15:         **if**  $\mathcal{U}_{nas} = \emptyset$  **then**
- 16:         **Finish.**
- 17:         **end if**
- 18:       **end if**
- 19:     **end if**
- 20:   **end for**
- 21:   **if** number of associated users with  $n = C$  **then**
- 22:     Break the **for** loop.
- 23:   **end if**
- 24:   **end if**
- 25:   **end for**
- 26: **end while**
- 27: **return**  $n$

---

all AAVs. We make this assumption to determine the upper bound of our solution's performance. Under this assumption in practice, each user must measure CSI from all AAVs and report it back to the MBS, which could require considerable overhead. The proposed algorithm, named AAV deployment and positioning algorithm (UDPA), selects the AAVs to be deployed and chooses their positions. The pseudo-code of UDPA is given in Algorithm 1 where the following notations are used:  $\mathcal{U}_{nas}$  is the set of not-yet associated users,  $\mathcal{Q}$  is the set of possible positions,  $p_{BS}$  is the available power at the MBS and  $p_n$  is the available power at AAV  $n$ .

The main idea of UDPA is to deploy AAVs one-by-one while selecting the best position from  $\mathcal{Q}$  until all (P1) constraints are satisfied. In line 2, while there is not-yet assigned users, UDPA adds one more AAV, then the **for** loop (line 4) selects its best position. In line 5, UDPA allocates the minimum transmit power levels to the deployed AAVs while satisfying the outage probability requirements, i.e., constraints (P1d) and (P1i). In line 8, UDPA sorts the not-yet-associated users in  $\mathcal{U}_{nas}$  according to their decreasing channel gains  $g_{u,n}$ . Then, the association procedure starts by iterating the available channels and users in lines 9–20. If the deployed AAVs can not satisfy all (P1) constraints,

**Algorithm 2** GA Algorithm

---

```

1: Randomly initialize the population  $N_{ind}$ .
2: Determine the fitness  $fitness[k]$  of each individual  $k$  in
    $N_{ind}$ .
3: for each generation do
4:   Select the best parents for the next population.
5:   Crossover the parents to create new individuals.
6:   if The constraints of (P1) are not respected then
7:     perform a correction for the new individuals.
8:   end if
9:   Mutate the selected individuals
10:  if The constraints of (P1) are not respected then
11:    perform a correction for the new individuals.
12:  end if
13:  Calculate the fitness of the new population.
14: end for
15: return Best individual in the last population.

```

---

more AAVs will be deployed. After each association,  $\mathcal{U}_{nas}$ ,  $p_{BS}$ , and  $p_n$  are updated. UDPA halts when all users are satisfied. The worst-case computational complexity of UDPA is  $\mathcal{O}(U|Q|(U \log U + UC^2))$ .

**V. GENETIC SOLUTION**

GA is a search-based optimization technique based on the principles of biological evolution. It is used to find optimal or near-optimal solutions to complex problems that are too difficult to solve with traditional methods. GA makes use of the AAV positions obtained from the heuristic algorithm (UDPA); hence, GA does not optimize the positions of AAVs. GA repeatedly modifies a population of individuals representing possible solutions and creates a new generation at each step. Over successive generations, the population evolves toward an optimal solution. In our implementation, an individual representing a solution is composed of 7 genes that are related to the optimization parameters ( $\mathbf{Z}$ ,  $\mathbf{Q}$ ,  $\mathcal{P}$ ,  $\mathbf{b}$ ). First, an initial random population of individuals, representing feasible solutions, is created. These individuals form the starting point for the evolutionary process. The fitness of each individual  $k$ , defined as  $fitness[k] = 1/\sum_{n=1}^N b_n$ , is evaluated after crossover and mutation operations. The next generation is formed from the actual generation undergoing the genetic operations: elitism selection, crossover, and mutation. The pseudo-code of GA is given in Algorithm 2.

- *Elitism selection*: According to a selection ratio  $\gamma_e$ , the individuals with the best fitness are chosen to be part of the next generation.
- *Crossover*: The parents for a potential crossover are selected by the roulette wheel method.
- *Mutation*: The mutation operation is controlled by a probability  $\gamma_m$ . An individual is mutated by changing a value randomly in the gene  $Z$ . The mutated individual replaces the original one in the next generation if it respects (P1) constraints.

A correction method is applied following each crossover and mutation operations to accelerate convergence and to ensure that the constraints in (P1) are respected. The correction after crossover is proceeded as follows. First, the algorithm verifies whether the associated users can maintain their associations even after the positions of the AAVs are changed. In cases where the QoS and power constraints can not be satisfied, the algorithm proceeds to search for an available channel among the already deployed AAVs to associate the disassociated users with respect to the constraints in (P1). If all the users are associated and all the constraints are met, the newly formed individual will be integrated into the generation. Otherwise, the parent is kept for the next generation. The correction after the mutation consists of (i) checking if the association obtained after the random change is feasible (the new channel chosen randomly is available), (ii) adjusting the allocated power level according to the new channel gain, and (iii) checking if the new association respects the outage probability thresholds. If all the constraints in (P1) are met, the mutated individual will replace the current individual in the next generation. The algorithm ends when the number of maximum generations is reached. The worst case complexity of this algorithm is  $\mathcal{O}(N_{gen}N_{ind}(\log N_{ind} + (U + UNC) + UN) + N_{ind}O_{pop}^{init})$ , where  $N_{gen}$  is the number of generations,  $N_{ind}$  is the number of individuals in a population and  $O_{pop}^{init}$  is the complexity of generating an initial individual.

**VI. DEEP REINFORCEMENT LEARNING SOLUTION**

In this section, we present the DDPG-based AAVs positioning and resource allocation algorithm (DUPRA). The optimization problem is formulated as a Markov Decision Process (MDP) and is composed of five elements ( $\mathcal{S}$ ,  $\mathcal{A}$ ,  $\mathcal{R}$ ,  $P$ ,  $\pi$ ). As this work minimizes the number of deployed AAVs, multi-agent solution where AAVs act as individual agents introduces complexity and convergence challenges. Moreover, choosing users as agents would require each user to know the CSI of all the communication links in the system. This assumption is both impractical and unrealistic, as it is not feasible for users to control or determine the positions of the serving AAVs. Thus, we consider a centralized BS to determine the deployed AAVs and their positions. Specifically, at each decision time step  $t$ , the MBS, which is considered as the central agent, observes state  $s(t)$  in the state space  $\mathcal{S}$  and chooses an action  $a(t)$  from the action space  $\mathcal{A}$  on the basis of its policy  $\pi$ . The agent receives the immediate reward  $r(t)$  by interacting with the environment and the experience tuple  $(s(t), a(t), r(t), s(t+1))$  is stored in the replay buffer memory for training.

- *State space  $\mathcal{S}$* : We define the system state at time step  $t$  as  $s(t) \in \mathcal{S}$ , where  $s(t)$  is the number of deployed AAVs at time step  $t - 1$ .
- *Action space  $\mathcal{A}$* : The agent's action consists of placing the AAVs, where the action at time step  $t$  is  $a(t) = Q(t)$  and  $Q(t)$  represents the locations of the AAVs at time step  $t$ .

- **Reward  $\mathcal{R}$ :** The reward is obtained according to the chosen action by the agent. The objective is to minimize the number of deployed AAVs. Furthermore, to ensure that all users are satisfied by a serving AAV at each time step  $t$ , the coverage constraint of the users should be taken into account in the reward function. If certain users are not satisfied by any AAV, a penalty  $c < 0$  is incurred by MBS. The instantaneous reward is calculated as follows:

$$r(t) = \begin{cases} \frac{1}{\sum_{n=1}^N b_n^{(t)}} & \text{if all users are satisfied} \\ c & \text{otherwise.} \end{cases} \quad (17)$$

### A. PRELIMINARIES

We first provide a brief introduction to DDPG (Deep Deterministic Policy Gradient), which is a DRL algorithm based on the actor-critic framework [46]. In DDPG, the critic  $Q(s, a; \theta^Q)$  evaluates the action-value function under the actor policy  $\mu(s; \theta^\mu)$ , where  $\theta^\mu$  and  $\theta^Q$  refer to the parameters of the actor and the critic networks.

However, a non-linear function approximator, e.g., deep neural networks (DNNs), is known to be unstable and can even cause divergence when applied in DRL. Thus, two techniques are usually used in DRL to prevent this issue: experience replay and target network [47]. DRL samples a mini-batch of experiences from the replay buffer. The random samples break the correlation between successive samples and stabilize the training process. Furthermore, the target networks of actor and critic  $\mu'(s; \theta^{\mu'})$  and  $Q'(s, a; \theta^{Q'})$  have the same architectures as the main networks and need to be updated by slowly tracking the main networks in order to train the critic network. The latter can be trained by minimizing the loss which is expressed as:

$$L(\theta^Q) = \frac{1}{B} \sum_t \left[ y_t(t) - Q(s(t), a(t); \theta^Q) \right]^2, \quad (18)$$

where  $y_t(t)$  is the update target and is given by:

$$y_t(t) = r(t) + \gamma Q'(s(t+1), \mu'(s(t+1); \theta^{\mu'}); \theta^{Q'}). \quad (19)$$

The discount factor  $\gamma$  is used for updating and  $B$  is the mini-batch sampling size. In addition, the actor is trained by minimizing the actor loss which is calculated as follows:

$$L(\theta^\mu) = \frac{1}{B} \sum_t -Q(s(t), \mu(s(t+1); \theta^\mu); \theta^Q). \quad (20)$$

Let  $\varepsilon \ll 1$  be the update rate, the soft update parameters of the target networks are given by [48]:

$$\theta' \leftarrow \varepsilon \theta + (1 - \varepsilon) \theta'. \quad (21)$$

### B. DUPRA ALGORITHM

DUPRA is detailed in Algorithm 3. At the beginning of each episode, the environment is reset, and no AAV is deployed in the initial state  $s(0)$ .

At each time step  $t$ , the MBS chooses an action  $a(t)$ , through the action network  $\mu(s(t); \theta^\mu)$ , adding an exploration noise  $N_e$  in line 6, in order to promote exploration and prevent the agent from falling into a local optimization during training. We choose uncorrelated mean-zero Gaussian noise, with the standard deviation decreasing over time steps. The noise variance is initially set to  $N_e^{max}$ . For a given number of time steps, the exploration noise is canceled and the agent is no longer exploring but directly selecting actions. Based on the selected action, the agent deploys the required AAVs to serve all users and assigns the minimum transmission power levels using the approach outlined in Algorithm 1. This involves deploying the AAVs one by one to their determined positions until all (P1) constraints are satisfied. If the MBS succeeds in satisfying all the users, then it receives a positive reward as described above and obtains the next state  $s(t+1)$ . Otherwise, the agent receives a negative reward, and the next state is set to 0. Then, the agent stores the corresponding transition tuple  $(s(t), s(t), s(t+1), r(t))$  in the experience replay buffer.

The update of the main networks is performed in lines 11–12. At each time step, if the buffer is full, a mini-batch of experiences is sampled from the replay buffer. This sample is used to train and update the networks by minimizing the actor and the critic losses. Then the target networks are updated by tracking the main networks in line 13.

### C. COMPLEXITY ANALYSIS OF DUPRA

To analyze the computational complexity of DUPRA, we define  $N_l^a$  as the number of neurons in the  $l$ th layer of the actor network and  $N_k^c$  as the number of neurons in the  $k$ th layer of the critic network. The critic and actor networks are fully connected networks. The computational complexity of the actor network is  $\mathcal{O}(\sum_{l=2}^{L-1} N_{l-1}^a N_l^a + N_l^a N_{l+1}^a)$ , where  $L$  is the number of layers of the actor network. Similarly, the computational complexity of the critic network is  $\mathcal{O}(\sum_{k=2}^{K-1} N_{k-1}^c N_k^c + N_k^c N_{k+1}^c)$ , where  $K$  is the number of layers of the critic network.

The actor and critic networks are simultaneously trained by extracting  $B$  experiences from the replay buffer for backpropagation training. Therefore, the complexity of Algorithm 3 is  $\mathcal{O}(TB(\sum_{k=2}^{K-1} N_{k-1}^c N_k^c + N_k^c N_{k+1}^c + \sum_{l=2}^{L-1} N_{l-1}^a N_l^a + N_l^a N_{l+1}^a))$ , where  $T$  is the number of time steps in each episode.

## VII. SIMULATION RESULTS

This section presents Monte-Carlo simulation results to evaluate the performance of the proposed solutions. We consider a circle area with a radius  $R = 50\text{m}$ . The MBS is placed in the center of the serving zone, whereas the users (remotely controlled robots) are randomly and uniformly located in the area. The genetic algorithm parameters (i.e., the elite selection ratio, mutation, and crossover probabilities) are empirically selected by simulation. The DDPG hyper-parameters (e.g., number of layers, number of neurons in

**Algorithm 3** DUPRA Algorithm

- 1: Randomly initialize the actor network  $\mu(s; \theta^\mu)$ , the target actor network  $\mu'(s; \theta^{\mu'})$  with weights  $\theta^\mu = \theta^{\mu'}$ , the critic network  $Q(s, a; \theta^Q)$ , and the target critic network  $Q'(s, a; \theta^{Q'})$  with weights  $\theta^Q = \theta^{Q'}$ .
- 2: Initialize the replay buffer.
- 3: **for** each episode **do**
- 4:   Reset the environment.
- 5:   **for** each time step  $t$  **do**
- 6:     Choose an action  $a(t) = \mu(s(t); \theta^\mu) + N_e$ .
- 7:     Allocate power levels and associate the users.
- 8:     Calculate  $r(t)$  as in (17) and move to  $s(t+1)$ .
- 9:     Store the tuple  $(s(t), a(t), s(t+1), r(t))$ .
- 10:     Sample a mini-batch of  $B$  transitions.
- 11:     Update  $\theta^Q$  by minimizing the critic loss as in (18).
- 12:     Update  $\theta^\mu$  by minimizing the actor loss as in (20).
- 13:     Update for the target networks:  $\theta^{Q'} = \varepsilon\theta^Q + (1 - \varepsilon)\theta^{Q'}$  and  $\theta^{\mu'} = \varepsilon\theta^\mu + (1 - \varepsilon)\theta^{\mu'}$ .
- 14:   **end for**
- 15: **end for**

**TABLE 3.** Notations and parameters values.

Notations	Parameters	Default values
$H_{uav}$	Height at which UAVs fly	30 m
$H_{bs}$	Height of the MBS antenna	2 m
$C = C_{ul} = C_{dl}$	Number of front-haul channels	4
$\alpha, \beta, \sigma_\xi$	LoS Path-loss parameters (f=28GHz)	61.4, 2, 5.8 [42]
	NLoS Path-loss parameters (f=28GHz)	72, 2.92, 8.7 [42]
$m_{LoS}$	<i>Nakagami</i> parameter for LoS links	3 [49]
$m_{NLoS}$	<i>Nakagami</i> parameter for NLoS links	2 [49]
$\sigma^2$	AWGN power spectral density	-174 dBm/Hz
$B_{fh}$	Available bandwidth for front-haul links	1 GHz
$B_{bh}$	Available bandwidth for back-haul links	10 GHz
$\zeta_{fh}$	UL and DL front-haul data rate threshold	500 Mbps
$\zeta_{bh}$	UL and DL back-haul data rate threshold	$C \times \zeta_{fh}$
$p_n^{max}$	Maximum power of the $n$ th UAV	30 dBm
$P_{BS}^{max}$	Maximum power of the MBS	40 dBm
$p_u^{max}$	Maximum power of each user	30 dBm
$\theta$	Width of transmitting antenna	90°
$G_m, G_s$	Main-lobe and side-lobe antenna gain	12 dB, 0 dB
$\gamma_e$	Elite selection ratio	0.1
$\gamma_c$	Crossover probability	0.6
$\gamma_m$	Mutation probability	0.2
$N_{gen}$	Number of generations	500
$N_{ind}$	Number of individuals in a generation	50
$l_{ra}$	Actor learning rate	0.0003
$l_{rc}$	Critic learning rate	0.0003
$\gamma$	Discount factor for update	0.99
$\varepsilon$	Soft update rate	0.005
$M$	Buffer size	1024
$B$	Sample batch size	64
$T$	Time step in an episode	600
$(N_1^a, N_2^a, N_3^a)$	Number of neurons in actor network	(1024, 512, 256)
$(N_1^c, N_2^c, N_3^c)$	Number of neurons in critic network	(1024, 512, 256)

both actor and critic network, learning rate) are also selected by simulations. The discount factor and the soft update rate values are chosen similarly to [48]. All the simulation parameters and their default values are listed in Table 3. Unless otherwise mentioned, these default values are used in the following simulations.

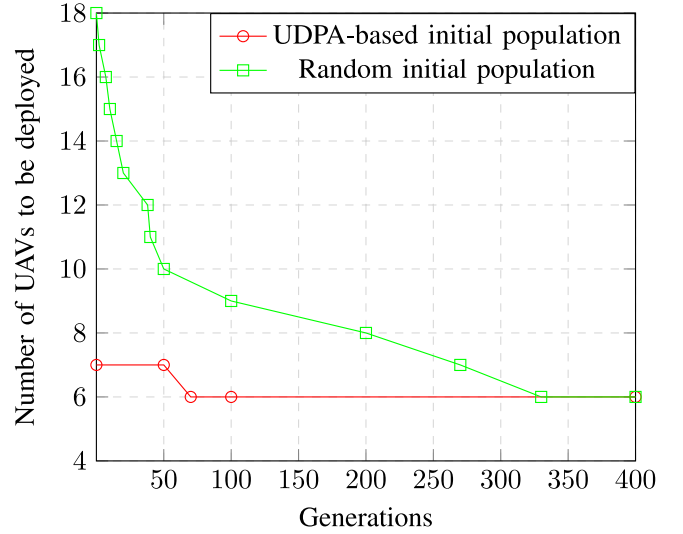
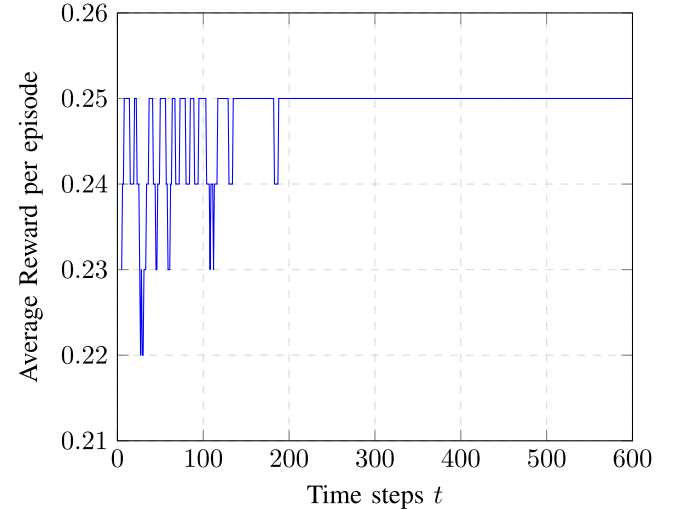
**FIGURE 2.** Convergence of the GA solution, for  $\zeta_m = 700$  Mbps,  $\hat{O}_m = \hat{O}_{bh} = 5\%$  and  $U = 18$ .**FIGURE 3.** DUPRA convergence, for  $U = 14$ .

Fig. 2 plots the convergence of GA when the initial population is randomly chosen and when the initial solution is obtained by greedy UDPA. As expected, using UDPA to obtain an initial population yields faster convergence. According to [50], GA converges to a global optimal solution as the number of generations increases.

To evaluate the performance of our DRL-based solution, DUPRA, we choose the average reward per episode as a metric. Fig. 3 shows the convergence of DUPRA for 14 users. It is clear that the average reward converges fast (i.e., after 200 time steps). Importantly, the agent, which is the MBS, is trained offline in a simulated environment, meaning that the AAVs are not included in the training process and, thus, their energy consumption is not impacted by the training. Moreover, changing the environment in each episode allows us to have more realizations of the environment that the DRL can observe in a real case. While

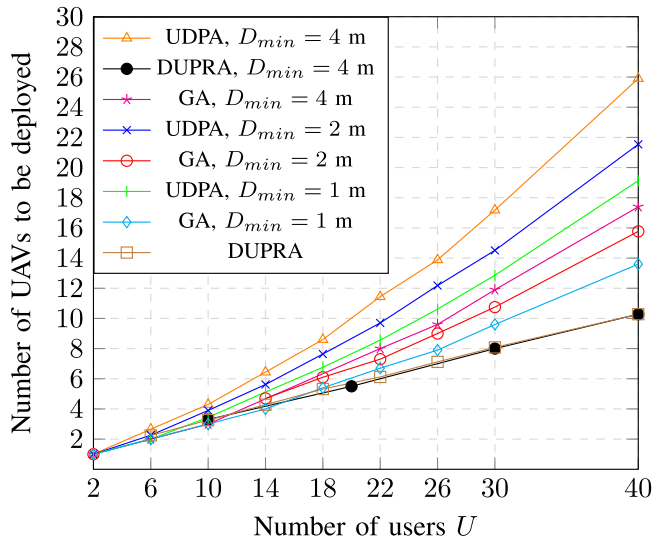


FIGURE 4. Impact of minimal distance between AAVs possible positions  $D_{min}$ , for  $\zeta_{fh} = 700$  Mbps and  $\hat{O}_{bh} = \hat{O}_{bh} = 5\%$ .

TABLE 4. Algorithms' complexity for  $U = 30$  and  $D_{min} = 2$  m.

Algorithm	Order of Complexity (operations)
UDPA	$19.4 \times 10^5$
GA	$2.3 \times 10^9$
DUPRA	$7.8 \times 10^{10}$

DRL shows significant potential for particular tasks and environments, its ability to generalize is limited without the incorporation of additional techniques and strategies.

In order to evaluate the impact of the discretization of the space on the performance of GA and UDPA algorithms, Fig. 4 compares these algorithms to DUPRA for different values of  $D_{min}$ , the minimal distance between two discretized possible positions of AAVs. It is to be reminded that DUPRA considers a continuous space. We can observe that the genetic algorithm (i.e., GA) outperforms the greedy one (i.e., UDPA) by a gap of approximately 25% but with higher complexity, as shown in Table 4. DUPRA results are close to GA for a small number of users but the performance gap becomes more important for a larger number of users. DUPRA results with  $D_{min} = 4$  m are derived from those of DUPRA where each value is approximated to the nearest position from the set of positions for  $D_{min} = 4$  m. This is because GA does not optimize the positions of AAVs; instead, it utilizes the resulting positions from UDPA. We observe that when  $D_{min}$  is smaller, the AAVs dispose of a larger set of positions to choose from. Thus, the number of AAVs to be deployed decreases. On the other side, the algorithms find only the positions of the AAVs and do not investigate the trajectory. The algorithms do not define how each AAV gets to its chosen position. Hence, avoiding collisions is not a task of our proposed algorithms, and it is out of the scope of this work. Nevertheless, the constraint (P1n) ensures that two AAVs can not have the same position. The proposed greedy and genetic algorithms and even the DRL-based

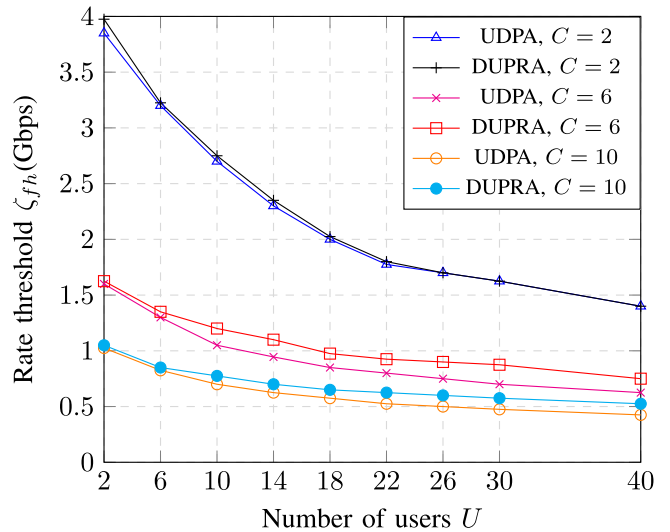


FIGURE 5. Feasibility region for  $\hat{O}_{fh} = \hat{O}_{bh} = 7\%$ .

solution are centralized. Centralized solutions generally face scalability challenges compared to distributed approaches. As the number of ground users increases, the problem's dimensionality grows accordingly. However, our centralized solutions can be adapted by dividing large networks into smaller ones and applying them within each smaller network.

Fig. 5 considers different values of available channels,  $C$ , and shows the feasibility region of the system for  $\hat{O}_{fh} = \hat{O}_{bh} = 7\%$ . The feasibility region is defined as the under-curve area gathering the set of possible rate thresholds satisfying the system constraints given the number of users. We can observe that the feasibility region is smaller when the number of available channels increases for both UDPA and DUPRA. It is to be noted that the feasibility regions for UDPA and GA are the same since greedy UDPA is used to generate the initial population for the proposed GA-based solution. For instance, considering  $C = 2$ , UDPA (and DUPRA) can no satisfy  $U = 22$  users when the rate threshold is higher than  $\zeta_{fh} = 1.8$  Gbps even if we deploy one AAV for each user. On the other hand, even when the number of available channels is larger,  $C = 6$ , UDPA can not satisfy  $U = 22$  users when the rate threshold is set to  $\zeta_{fh} = 0.8$  Gbps whereas DUPRA can satisfy  $U = 22$  users if the rate threshold is smaller than 0.9 Gbps. Clearly, DUPRA extends the feasibility region, compared to UDPA. This is due to the continuous positions that DUPRA chooses for the deployed AAVs. To conclude, increasing the number of available channels leads to a decrease in the bandwidth of each channel, which in turn results in reducing the feasibility region since it becomes more challenging to satisfy all the users.

Without loss of generality, the next figures (i.e., Fig. 6, 7 and 8) consider the greedy UDPA to evaluate the impact of the system parameters on the performance since UDPA has relatively low computational complexity. Fig. 6 shows the impact of  $\hat{O}_{fh}$  and  $\zeta_{fh}$  on the system performance, where

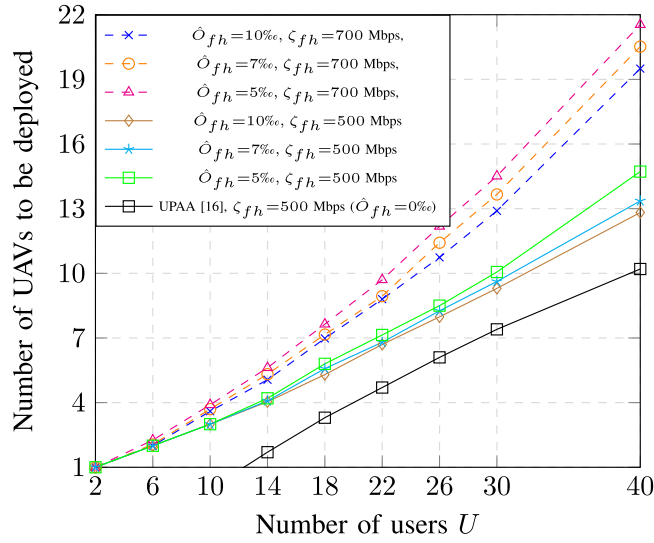
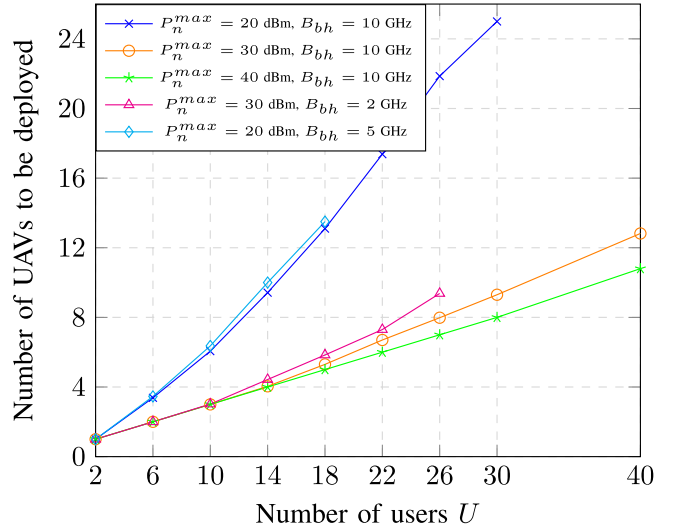


FIGURE 6. Performance of UDPA.

$\hat{O}_{fh} = \hat{O}_{bh}$ . Clearly, increasing  $\zeta_{fh}$  increases the required number of AAVs. Also, decreasing the outage probability threshold increases the number of AAVs to be deployed. For instance, decreasing  $\hat{O}_{fh}$  from 10% to 7% increases the minimum number of AAVs to be deployed by approximately 10 – 15%. We can also observe that, for a large number of users and an outage probability of 5%, the system needs to deploy on average more than one AAV for every two users to ensure a data rate of 700 Mbps. Meanwhile, the system needs to deploy  $\sim 30\%$  less AAVs to satisfy  $\zeta_{fh} = 500$  Mbps. This is due to the stringent QoS requirements and the interference caused by other transmissions. Fig. 6 also compares the performance of our heuristic UDPA with UPAA proposed in [16]. As expected, the number of AAVs deployed by UPAA is fewer than those deployed by UDPA. This can be explained by the fact that UPAA is able to associate users with the MBS whereas our system model supposes that it is not possible. It is also important to note that UPAA only satisfies rate constraints only on downlink links, whereas UDPA has to satisfy rate and outage constraints on both uplink and downlink links.

Fig. 7 shows both the impact of the available transmit power at the AAVs,  $p_n^{max}$ , and the impact of the back-haul bandwidth,  $B_{bh}$ , on the number of AAVs to be deployed. It considers  $\zeta_{fh} = 500$  Mbps and  $\hat{O}_{fh} = \hat{O}_{bh} = 10\%$ . Usually, the energy consumption of the AAV includes both the transmit power (or communication power) and the propulsion power. However, since our study focuses on a single time slot, only the transmission power was considered and we studied its impact on the performances of the proposed solution. It is to be noted that some curves are not completed because they fall out of the feasibility regions. It is shown that when  $p_n^{max}$  is low, (e.g.,  $p_n^{max} = 20$  dBm), the number of AAVs to be deployed increases rapidly with the number of users, whereas the difference in performance is small when  $p_n^{max}$  is higher. These curves show the huge impact of the


 FIGURE 7. Impact of the AAVs maximum transmit power and  $B_{bh}$ , for  $\hat{O}_m = \hat{O}_{bh} = 10\%$  and  $\zeta_m = 500$  Mbps.

maximum transmit power on the number of AAVs required to satisfy all the users. Regarding the impact of limited back-haul resources, we observe that for  $p_n^{max} = 30$  dBm and  $B_{bh} = 2$  GHz, satisfying all the users becomes impossible for  $U > 26$ . This maximum number of users that can be satisfied becomes 18 when  $p_n^{max} = 20$  dBm and  $B_{bh} = 5$  GHz. We can also observe that when the problem is feasible (i.e., all users can be satisfied), the number of AAVs to be deployed increases slightly when the available bandwidth at the back-haul is limited.

When the problem is infeasible, we propose to make use of a user selection method and thus the problem becomes to satisfy the maximum number of users. The users are selected according to their decreasing channel gains. Fig. 8 shows the performance of the proposed greedy UDPA solution using the proposed user selection method to satisfy the maximum number of users. We consider three scenarios with different parameter values. Scenario 1 considers  $P_n^{max} = 20$  dBm and  $B_{bh} = 10$  GHz, whereas scenario 2 considers  $P_n^{max} = 30$  dBm and  $B_{bh} = 2$  GHz, and scenario 3 we considers  $P_n^{max} = 20$  dBm and  $B_{bh} = 5$  GHz. The results show that for scenarios 1 and 2, the number of AAVs to be deployed reaches its maximum value when the number of users in the system is larger and the percentage of satisfied users is lower. Moreover, increasing the bandwidth in back-haul links but having a limited AAVs' transmit power (scenario 3), increases the number of deployed AAVs and decreases the number of satisfied users.

Fig. 9 shows the impact of the number of available front-haul channels,  $C = C_{ul} = C_{dl}$ , on the required number of AAVs considering a fixed  $B_{fh} = 1$  GHz. Note that some curves are not completed since they reach the infeasibility regions. Comparing the performance of UDPA for different values of  $C$  (while the total bandwidth  $B_{fh}$  is fixed), it is to be noticed that when  $C$  increases, and thus the channels become narrower, the required number of AAVs decreases

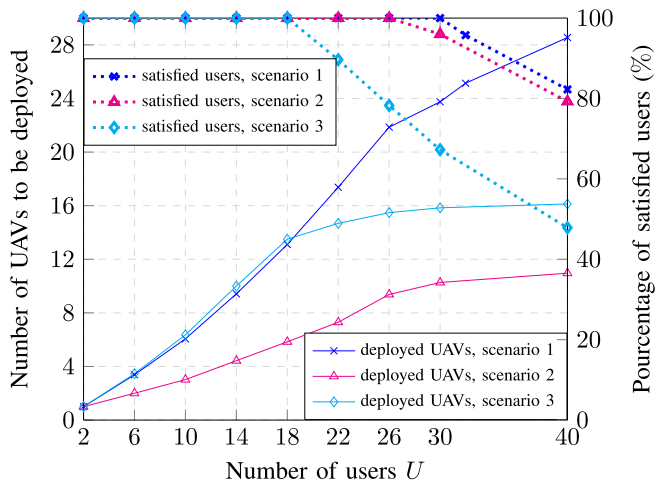


FIGURE 8. Performance with user selection with limited AAVs maximum transmit power and  $B_{bh}$ , for  $\hat{O}_m = \hat{O}_{bh} = 10\%$  and  $\zeta_{fh} = 500$  Mbps.

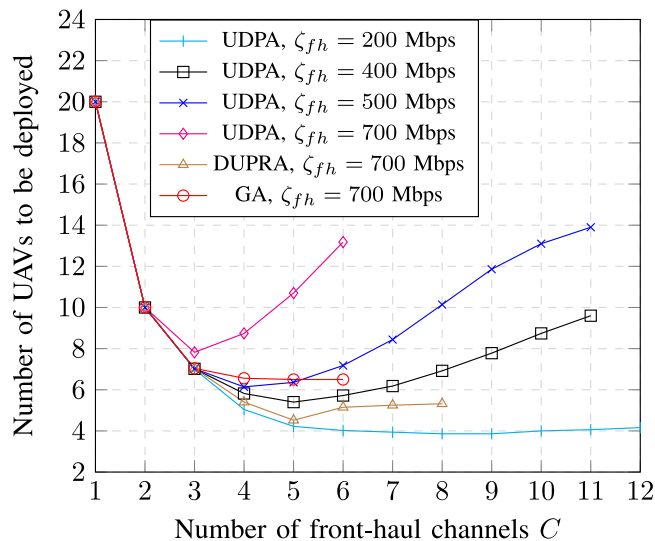


FIGURE 9. Impact of the number of front-haul channels, for  $U = 20$  and  $\hat{O}_m = \hat{O}_{bh} = 7\%$ .

and stabilizes for a while before increasing. This behavior can be explained by the fact that higher  $C$  could decrease the required number of AAVs since one AAV can serve more users, but at one point, the channels become too narrow. Hence, it becomes more difficult to satisfy the users' requirements and more AAV need to be deployed. Clearly, there is an optimal value of the number of front-haul channels to find. In our experiments, this value is 5, 4 and 3 channels for  $\zeta_{fh} = 300$  Mbps, 500 Mbps and 700 Mbps, respectively. When  $\zeta_{fh} = 200$ , UDPA deploys the minimum number of deployed AAVs, regardless of the number of available channels. However, the performances of DDPG and GA are shown to be superior to those of UDPA for  $\zeta_{fh} = 700$  Mbps. In fact, the number of AAVs required for a successful deployment is significantly smaller when using DUPRA and GA. Moreover, DDPG extends the feasibility of the problem with  $C = 8$  by enabling continuous positioning of the AAVs.

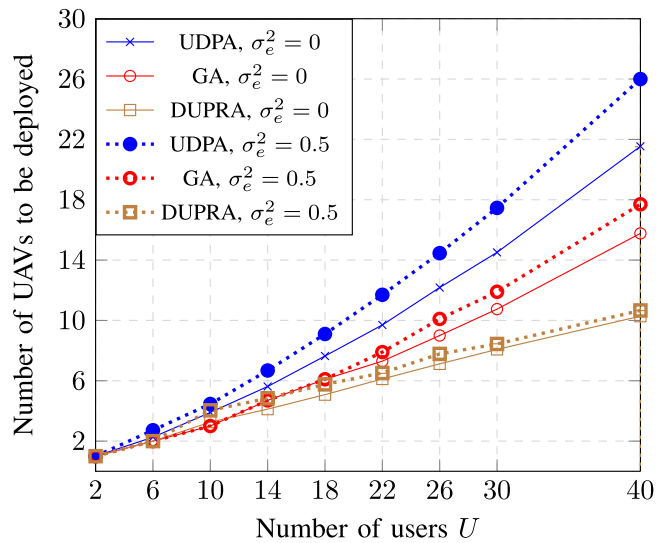


FIGURE 10. Impact of the Imperfect CSI, for  $\zeta_{fh} = 700$  Mbps and  $\hat{O}_m = \hat{O}_{bh} = 5\%$ .

Fig. 10 studies the performances of the proposed algorithms when considering imperfect knowledge of the channel state information. The variance of CSI estimation error is not measured by the system. The variance estimation is out of our scope. We suppose the variance to be known for evaluating the effect of imperfect CSI on the performances of our solutions. To assess its impact on the performance of our solutions, we make a simplifying assumption by assigning the same error variance across all links. While this assumption may not fully reflect realistic scenarios, it is adopted for simplicity's sake and to facilitate the evaluation of its impact on system performance. We observe, with an important error's variance  $\sigma_e^2 = 0.5$ , a significant deterioration of the proposed algorithm's performance using UDPA solution leading to increasing the number of deployed AAVs by 20% for a large number of users. As for DUPRA, we observe that its performance is more robust and the number of AAVs to be deployed is slightly increasing by 3.5% with imperfect CSI. GA is also robust for low number of users, but for a larger number of users, it deploys 12% more AAVs. The confidence interval is not presented in the figures for clarity purposes. Its bounds are within 1% of the mean value for the three algorithms. However, if the coherence time is shorter than the time slot, no solution can effectively handle the rapid real-time changes in channel conditions.

## VIII. CONCLUSION

This paper studied the joint problem of AAVs positioning and resource allocation in a AAV-assisted wireless network for stringent tele-operation applications. These applications introduce stringent requirements for UL and DL communications in terms of rate and outage probability thresholds. The tackled problem was modeled as a non-convex MINLP problem where the objective function is to minimize the number of deployed AAVs while satisfying

all users' requirements. Since the considered optimization problem is proved to be  $\mathcal{NP}$ -hard, we proposed a greedy solution, a GA-based solution, and a DRL-based solution, to solve it efficiently. Simulation results show that the performance of the greedy solution is close to that of the genetic-based and deep reinforcement methods, with a significant reduction in computational complexity. Additionally, the results demonstrate the effectiveness of the proposed algorithmic solutions in terms of the required number of deployed AAV to meet the data transmission requirements of all users in both UL and DL directions. In future work, we will further consider a complex scenario and propose decentralizing the considered multi-AAV system. We will propose a distributed DRL solution to resolve our resource allocation problem. This solution can potentially integrate multi-agent deep reinforcement learning to leverage an autonomous communication system. Moreover, we will consider the path loss models published in 3GPP Rel 17 and introduce a mobility model.

## REFERENCES

- [1] A. Aijaz, "Toward human-in-the-loop mobile networks: A radio resource allocation perspective on haptic communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4493–4508, Jul. 2018.
- [2] T. Taleb, Y. Hadjadj-Aoul, and T. Ahmed, "Challenges, opportunities, and solutions for converged satellite and terrestrial networks," *IEEE Wireless Commun.*, vol. 18, no. 1, pp. 46–52, Feb. 2011.
- [3] P. Wang, J. Zhang, X. Zhang, Z. Yan, B. G. Evans, and W. Wang, "Convergence of satellite and terrestrial networks: A comprehensive survey," *IEEE Access*, vol. 8, pp. 5550–5588, 2020.
- [4] S. Watts and O. G. Aliu, "5G resilient backhaul using integrated satellite networks," in *Proc. 7th Adv. Satellite Multimedia Syst. Conf. 13th Signal Process. Space Commun. Workshop (ASMS/SPSC)*, 2014, pp. 114–119.
- [5] A. Kapovits et al., "Satellite communications integration with terrestrial networks," *China Commun.*, vol. 15, no. 8, pp. 22–38, 2018.
- [6] T. Bouzid, N. CHAIB, M. Bensaad, and O. Oubbati, "5G network slicing with unmanned aerial vehicles: Taxonomy, survey, and future directions," *Trans. Emerg. Telecommun. Technol.*, vol. 34, no. 3, Dec. 2022, Art. no. e4721.
- [7] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1123–1152, 2nd Quart., 2016.
- [8] Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "Toward future unmanned aerial vehicle networks: Architecture, resource allocation and field experiments," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 94–99, Feb. 2019.
- [9] A. Fotouhi et al., "Survey on UAV cellular communications: Practical aspects, Standardization advancements, regulation, and security challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3417–3442, 4th Quart., 2019.
- [10] D.-H. Tran, V.-D. Nguyen, S. Chatzinotas, T. X. Vu, and B. Ottersten, "UAV relay-assisted emergency communications in IoT networks: Resource allocation and trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1621–1637, Mar. 2022.
- [11] G. Zhang, X. Ou, M. Cui, Q. Wu, S. Ma, and W. Chen, "Cooperative UAV enabled relaying systems: Joint trajectory and transmit power optimization," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 543–557, Mar. 2022.
- [12] G.-M. Kang, H. M. Kim, and O.-S. Shin, "Bi-directional power and trajectory control for UAV-assisted cellular systems," in *Proc. ICTC*, 2021, pp. 69–72.
- [13] Y. Guo, S. Yin, and J. Hao, "Joint placement and resources optimization for multi-user UAV-relaying systems with underlaid cellular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12374–12377, Oct. 2020.
- [14] W. Shi et al., "Joint UL/DL resource allocation for UAV-aided full-duplex NOMA communications," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8474–8487, Jun. 2021.
- [15] H. Zeng, X. Zhu, Y. Jiang, Z. Wei, S. Sun, and X. Xiong, "Toward UL-DL rate balancing: Joint resource allocation and hybrid-mode multiple access for UAV-BS-assisted communication systems," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2757–2771, Apr. 2022.
- [16] A. Ahmed, C. Chaieb, W. Ajib, H. Elbiaze, and R. Glitho, "AAVs optimization in cell-free aerial communication networks," *Comput. Commun.*, vol. 232, Feb. 2025, Art. no. 108041.
- [17] M. Z. Hassan, G. Kaddoum, and O. Akhrif, "Interference management in cellular-connected Internet of Drones networks with drone-pairing and uplink rate-splitting multiple access," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16060–16079, Sep. 2022.
- [18] V. Roberge, M. Tarbouchi, and G. Labonte, "Comparison of parallel genetic algorithm and particle swarm optimization for real-time UAV path planning," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 132–141, Feb. 2013.
- [19] T. Yang, Y. Hu, X. Yuan, and R. Mathar, "Genetic algorithm based UAV trajectory design in wireless power transfer systems," in *Proc. WCNC*, 2019, pp. 1–6.
- [20] A. H. Sawalmeh, N. S. Othman, H. Shakhatreh, and A. Khreishah, "Wireless coverage for mobile users in dynamic environments using UAV," *IEEE Access*, vol. 7, pp. 126376–126390, 2019.
- [21] Y. A. Sambo, P. V. Klaine, J. P. B. Nadas, and M. A. Imran, "Energy minimization UAV trajectory design for delay-tolerant emergency communication," in *Proc. ICC*, 2019, pp. 1–6.
- [22] C. Chaieb, F. Abdelkefi, and W. Ajib, "Deep reinforcement learning for resource allocation in multi-band and hybrid OMA-NOMA wireless networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 187–198, Jan. 2023.
- [23] A. Alwarafy, M. Abdallah, B. S. Çiftler, A. Al-Fuqaha, and M. Hamdi, "The frontiers of deep reinforcement learning for resource management in future wireless HetNets: Techniques, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 322–365, 2022.
- [24] R. Ding, F. Gao, and X. S. Shen, "3-D UAV trajectory design and frequency band allocation for energy-efficient and fair communication: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7796–7809, Dec. 2020.
- [25] Z. Chang, H. Deng, L. You, G. Min, S. Garg, and G. Kaddoum, "Trajectory design and resource allocation for multi-UAV networks: Deep reinforcement learning approaches," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 5, pp. 2940–2951, Sep./Oct. 2023.
- [26] J. Li, C. Zhou, J. Liu, M. Sheng, N. Zhao, and Y. Su, "Reinforcement learning-based resource allocation for coverage continuity in high dynamic UAV communication networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 848–860, Feb. 2024.
- [27] G. B. Tarekegn et al., "A Centralized multi-agent DRL-based trajectory control strategy for unmanned aerial vehicle-enabled wireless communications," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 1230–1241, 2024.
- [28] A. I. Ameur, O. S. Oubbati, A. Lakas, A. Rachedi, and M. B. Yagoubi, "Efficient vehicular data sharing using aerial P2P backbone," *IEEE Trans. Intell. Veh.*, early access, Jun. 13, 2024, doi: 10.1109/TIV.2024.3414140.
- [29] K. Messaoudi, O. S. Oubbati, A. Rachedi, and T. Bendouma, "UAV-UGV-based system for AoI minimization in IoT networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 4743–4748.
- [30] M. Z. Hassan, G. Kaddoum, and O. Akhrif, "Resource allocation for joint interference management and security enhancement in cellular-connected Internet-of-Drones networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 12869–12884, Dec. 2022.
- [31] C. Wang, D. Deng, L. Xu, and W. Wang, "Resource scheduling based on deep reinforcement learning in UAV assisted emergency communication networks," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3834–3848, Jun. 2022.
- [32] Z. Dai, Y. Zhang, W. Zhang, X. Luo, and Z. He, "A multi-agent collaborative environment learning method for UAV deployment and resource allocation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 8, pp. 120–130, 2022.
- [33] P. Luong, F. Gagnon, L.-N. Tran, and F. Labeau, "Deep reinforcement learning-based resource allocation in cooperative UAV-assisted wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7610–7625, Nov. 2021.

- [34] S. Yin and F. R. Yu, "Resource allocation and trajectory design in UAV-aided cellular networks based on Multiagent reinforcement learning," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2933–2943, Feb. 2022.
- [35] A. Ahmed, C. Chaieb, W. Ajib, H. Elbiaze, and R. Glitho, "Resource allocation and UAVs placement in cell-free wireless networks," in *Proc. GLOBECOM*, 2022, pp. 5995–6000.
- [36] L. Zhang et al., "A survey on 5G millimeter wave communications for UAV-assisted wireless networks," *IEEE Access*, vol. 7, pp. 117460–117504, 2019.
- [37] M. Hammami, C. Chaieb, W. Ajib, H. Elbiaze, and R. Glitho, "UAV-assisted wireless networks for stringent applications: Resource allocation and positioning," in *Proc. WCNC*, Mar. 2023, pp. 1–6.
- [38] A. A. Khuwaja, Y. Chen, and G. Zheng, "Effect of user mobility and channel fading on the outage performance of UAV communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 367–370, Mar. 2020.
- [39] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.
- [40] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier Millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sep. 2015.
- [41] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
- [42] M. R. Akdeniz et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [43] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2874–2885, Dec. 2017.
- [44] Y.-J. Chen, K.-M. Liao, and Y.-F. Chen, "End-to-end delay analysis in aerial-terrestrial heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1793–1806, Feb. 2021.
- [45] R. Li and H.-C. Huang, "A general  $k$ -level uncapacitated facility location problem," in *Proc. Adv. Intell. Comput. Theories Appl. Aspects Contemp. Intell. Comput. Techn.*, 2008, pp. 76–83.
- [46] V. Konda and J. Tsitsiklis, "Actor-critic algorithms," in *Proc. NeurIPS*, vol. 12, 1999, pp. 1–7.
- [47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [48] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in *Proc. ICLR*, 2016, p. 12.
- [49] K. Humadi, I. Trigui, W.-P. Zhu, and W. Ajib, "Coverage analysis of user-centric millimeter wave networks under dynamic base station clustering," in *Proc. ICC*, 2021, pp. 1–6.
- [50] R. F. Hartl and R. Belew, *A Global Convergence Proof for a Class of Genetic Algorithms*, Univ. Technol., Vienna, Austria, 1990.