MADRL-Based 3D Deployment and User Association of Cooperative mmWave Aerial Base Stations for Capacity Enhancement

ZHAO Yikun, ZHOU Fanqin, FENG Lei, LI Wenjing, and YU Peng

(*State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*)

Abstract — **Although millimeter-wave aerial base station (mAeBS) gains rich wireless capacity, it is technically difficult for deploying several mAeBSs to solve the surge of data traffic in hotspots when considering the amount of interference from neighboring mAeBS. This paper introduces coordinated multiple points transmission (CoMP) into the mAeBS-assisted network for capacity enhancement and designs a two-timescale approach for three-dimensional (3D) deployment and user association of cooperative mAeBSs. Specially, an affinity propagation clustering based mAeBS-user cooperative association scheme is conducted on a large timescale followed by modeling the capacity evaluation, and a deployment algorithm based on multi-agent (MA) deep deterministic policy gradient (MADDPG) is designed on the small timescale to obtain the 3D position of mAeBS in a distributed manner. Simulation results show that the proposed approach has significant throughput gains over conventional schemes without CoMP, and the MADDPG is more efficient than centralized deep reinforcement learning (DRL) algorithms in deriving the solution.**

Key words — **Aerial base station, mmWave, Capacity enhancement, Cooperative communication, Multi-agent deep reinforcement learning (MADRL).**

I. Introduction

The increasing high-data-rate network services exert strong data traffic load pressure in user-dense areas, such as in big stadiums and concerts, which form hotspots and impact the consistency of service quality. Millimeter-wave (mmWave) communication has been one of the important technical means to improve the network capacity by exploring new frequency bands [1]. Although mmWave communication system has the ad- vantages of broad bandwidth, high antenna gain, and tiny component size, it is also faced with the challenges of high propagation loss, poor diffraction ability, etc. [2]. Especially in high-rise building intensive areas, the transmission of mmWave signal is probably blocked by obstacles, which seriously deteriorates the communication performance. The unmanned aerial vehicle (UAV)-assisted network typically performs the line-of-sight (LoS) communication with ground user equipments (UEs), which is very suitable for mmWave to achieve high beamforming gain [3]. mmWave aerial base station (mAeBS), mmWave BS mounted on UAV, has unique advantages, which can be concluded as [2]–[5]: 1) mAeBS can provide significant LoS links and reduce the blocking effect in the transmission process; 2) mAeBS can flexibly adjust its location with the change of distribution and traffic demand of UEs; 3) mmWave spectrum can provide higher capacity without interference to the current terrestrial networks; 4) The tiny component sizes of mmWave BS make it easy to be equipped on the space-limited UAV. In light of these benefits, mAeBS is expected to become an effective supplement to the existing cellular networks to address the needs of ultra-intensive services in hotspot areas.

Generally, a single mAeBS only serves a limited number of UEs. In order to expand the coverage area, multiple mAeBSs can be deployed together [6]. However, the interference between these mAeBSs becomes a critical impacting factor, since strong LoS links usually dominate in air-to-ground (A2G) channel and the cross-link interference is unavoidable. In the meanwhile, although the main lobe of mmWave beam has high directivity, the negative effect of the side lobe can

not be ignored because it will cause interference to the adjacent region [7]. Especially the number of antenna elements carried by UAV is limited due to the limited space of UAV, which results in a poor sidelobe suppression effect. Coordinated multiple points transmission (CoMP) has been viewed as a powerful interference control approach in traditional cellular systems [8]. Compared with other interference mitigation methods such as power control and frequency reuse, CoMP has the unique advantage of significantly improving the throughput performance since it can exploit the interference as a beneficial signal. In CoMP, each UE is served by several BSs collaboratively, and the interference signal is transformed into desired signal sources [9]. However, according to our knowledge, there is no previous work on CoMP-assisted multi-mAeBS networking for capacity enhancement in hotspot areas.

How to deploy multiple AeBSs collaboratively is also a challenging problem. Conventional optimization algorithms are usually impractical for problems in complex UAV-assisted networks, since they usually have high computational complexity, and they need to calculate again and again even when encountering a similar decision scenario. Deep reinforcement learning (DRL) provides a solution for decision systems to learn and use experience in changing environments [10]. The existing works have investigated the 3D deployment problems in AeBS networks with the help of DRL algorithms, such as Q-learning [11], deep Q-network (DQN) [12], dueling DQN (DDQN) [13]. However, these approaches follow a centralized paradigm, which perform poorly in scalability and flexibility due to their huge state and action space. Fortunately, these challenges can be handled by invoking multi-agent deep reinforcement learning (MADRL) algorithms, which conducts in distributed manners [14]. However, MADRL is inherently more complicated than single-agent DRL, since agents need to interact with other agents and the environment at the same time. It is still questionable whether it can be applied in UAV-assisted networks.

To deal with these challenges, this paper introduces CoMP into the multi-mAeBS network for capacity enhancement in hotspot areas, and designs a two-timescale mechanism for 3D deployment of mAeBSs and the mAeBS-UE association. The main contributions of this paper can be summarized as follows: 1) A capacity evaluation model of cooperative mAeBSs for wireless capacity enhancement in hotspot areas is established, considering the characteristics of A2G propagation, CoMP, and 3D mmWave beam scheduling. 2) On the large timescale, a clustering-based user association algorithm is proposed, which can divide mAeBSs that severely interfere with each other into a cluster, to form

cooperative mAeBS-UE associations, so as to achieve considerable capacity enhancement. 3) On the small timescale, the multi-agent deep deterministic policy gradient (MADDPG) algorithm is adopted to let each mAeBS make its motion decision based on its own observation, in the direction of maximizing the wireless network capacity. Particularly, an autonomous deployment management architecture is designed, in which mAeBSs exchange information with aerial base station controller (ABC) in the network side, without the need to communicate with each other directly.

The rest of the paper is organized as follows. Section II presents the related works and the system model is depicted in Section III. Then, the two-timescale mechanism for UE association and 3D deployment of mAeBSs is specified in Section IV. Section V presents the discussions of simulation results. Finally, Section VI makes a conclusion.

II. Related Work

There are some instructive survey papers on UAV mmWave communications [2]–[4], [15]. The first work on enabling AeBS with mmWave networks was explored in [15], where the key challenges and possible solutions in mmWave UAV networks are discussed. Several works have been studied focusing on A2G mmWave channel characteristics [16], [17], beam management [18]–[20], secure transmission [21] and deployment [6]. Early research on mAeBS networks mainly considers the scenario that only one mAeBS is deployed. Recently, with the increasingly intensive distribution of UEs, it is difficult to deploy a single mAeBS to meet the communication demands. Toward this end, the multi-mAeBS network calls for great research efforts.

1. Interference issue in the multi-AeBS networks

Interference is usually dominant in multi-UAVs networks and it severely deteriorates the communication performance. Many previous works adopt user clustering methods, such as K-means [12], [22] and its improved versions [23], [24], to cluster AeBSs and UEs based on geographical information, ignoring the inter-cluster interference to simplify the model, which is not the case in practical scenarios. Some works take interference into consideration. In [25], the minimum achievable system throughput is maximized in the presence of co-channel interference. Authors in [26] optimizes trajectory and power control of AeBSs to obtain a higher system sum rate, considering the cross-link interference. These works consider the interference in multi-AeBS networks, but they do not utilize CoMP to convert the interference into capacity gain.

CoMP has been widely adopted as an effective interference control technique. Extensive works have shown that it can improve the coverage and capacity performance in terrestrial cellular systems [8], [9]. Affinity propagation clustering (APC), first proposed in [27], has been used as a CoMP clustering method and has shown its good performance in improving the capacity performance [28]. In [28] and [29], APC is used in CoMP clustering to mitigate cell edge users' interference in heterogeneous cloud small cell networks and picocell systems, respectively. The above works mainly focus on the terrestrial network. In [30], power control and APC-based CoMP scheme are designed to reduce the interference in the UAV-assisted network. However, this paper assumes the network environment is static and does not consider the deployment design, which fails to fully exploit the mobility of UAVs.

2. Deployment design of multiple AeBSs

The 3D UAV deployment is one of the fundamental problems of UAV-assisted networks, and the adjustable altitude of UAVs and their potential mobility provide freedom for effective deployment. In [6], the authors decouple the joint optimization problem of the deployment, UE association and hybrid beamforming of mAeBSs into two subproblems and solve them alternately. Ref.[31] designs an iterative algorithm to jointly optimize AeBSs' deployment and UE association. Ref.[32] focuses on maximizing the throughput and proposes a heuristic algorithm to optimize the placement of AeBSs. These aforementioned algorithms usually have high complexity, and more importantly, and the results need to be calculated again once the network environment changes. The recalculation process is too slow for real-time operation, which greatly limits the application of these algorithms in practical multi-AeBS networks.

3. DRL in the deployment design of AeBSs

To address these issues, many works have invoked DRL algorithms to solve the deployment design problem. Authors in [11] use Q-Learning to make motion decisions for AeBSs to maximize the data which is collected from UEs while minimizing power consumption. In [12], a DQN-based mAeBS placement algorithm is designed to achieve maximal spectral efficiency considering the quality of service (QoS) of UEs. Ref.[13] uses DDQN to optimize the deployment of AeBSs with the movement of UEs for better downlink capacity performance. However, these above works assume that each UAV has complete network information, while in practical applications, due to the high movement speed of UAVs, it is difficult for each UAV to obtain a global understanding of the dynamic environment. At the same time, these aforementioned DRL algorithms are performed in a centralized way, and the size of action-space and state-space will explosively grow as the number of network nodes increases, which occupies a large amount of memory and brings a lot of time overhead. Additionally, these centralized DRL algorithms also face the problem of poor scalability and flexibility.

The emergence of MADRL provides a new solution to address these challenges. MADDPG, first proposed in [33], is a popular MADRL algorithm which has a good performance in multi-agent collaborative, competitive, or mixed environments, and has been used to manage the trajectory of UAVs recently. Ref.[34] invokes MADDPG algorithm to design the dynamic trajectory of each AeBS, aiming to maximize secure rate in the presence of ground eavesdroppers. Authors in [35] use MADDPG to manage the trajectory of each UAV independently on the premise of meeting the constraints of fairness and energy consumption. In [36], a trajectory design algorithm based on MADDPG is proposed to ensure the freshness of the collected data. These aforementioned articles are not in the scope of capacity enhancement, and do not consider mmWave characteristics and interference between aerial base stations, so they cannot be directly applied to this scenario. However, these articles show that MADDPG can effectively improve training efficiency, these factors inspire us to solve the deployment problem of mAeBSs by invoking this algorithm.

Motivated by these aforementioned works, we utilize an APC-based CoMP clustering scheme to address the interference issue and adopt the MADDPG to optimize the 3D deployment of mAeBSs for better throughput. Table 1 summarizes the difference between our work and the existing literature on the deployment of AeBSs.

III. System Model

We consider a multi-mAeBS downlink communication scenario where M mAeBSs serve U ground UEs, as shown in Fig.1 The set of mAeBS is denoted as $\boldsymbol{M} = \{1, 2, \ldots, M\}$, and the ground UEs set is represented as $\boldsymbol{U} = \{1, 2, \ldots, U\}$. The coordinates of mAeBS m and UE u are denoted as (x_m, y_m, h_m) and $(x_u, y_u, 0)$, respectively.

1. Air-to-ground channel model

In the A2G channel, the radio signal transmitted by mAeBS first propagates in free space, and then is shadowed and scattered by obstacles such as buildings and trees when it reaches the urban environment. This process brings excessive pathloss on top of the free space pathloss. Thus, the A2G pathloss between mAeBS m and ground user u under the LoS and non-line-of-sight (NLoS) propagation environment can be modeled as [38]:

Table 1. Comparison between our work and the existing literature on deployment of AeBSs

| Reference	Optimization goal	mmWave	Interference modeling	CoMP	Deployment method
[6]	Capacity	✓	✓	×	Alternating optimization
[11]	Data collection	×	×	×	Q-learning
[12]	Capacity	✓	×	×	DQN
[13]	Capacity	×	✓	×	DDQN
[14]	Capacity	×	✓	×	Multi-agent Q-learning
[23]	Transmission power	×	✓	×	Alternating optimization
[25]	Capacity	×	✓	×	Successive convex optimization
[26]	Capacity	×	✓	×	Successive convex optimization
[31]	Capacity	✓	✓	×	Iterative algorithm
[32]	Capacity	×	×	×	Heuristic algorithm
[34]	Security	×	×	×	MADDPG
[35]	Fairness	×	×	×	MADDPG
[36]	Data collection	×	✓	×	MADDPG
[37]	Security	×	✓	✓	Alternating optimization
Our work	Capacity	✓	✓	✓	MADDPG
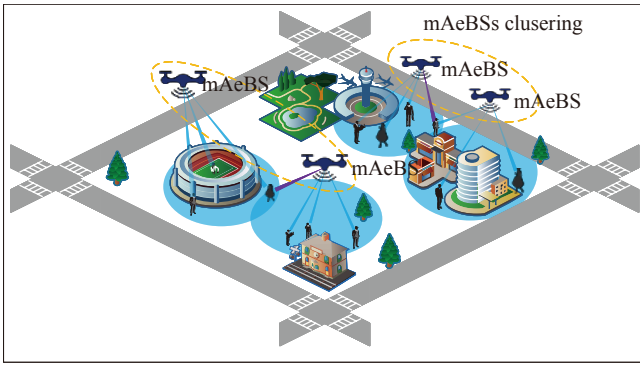
Fig. 1. Multi-mAeBS serving for hotspot area.

$$L_{LoS}^{m,u} = L_{FS}^{m,u} + \eta_{LoS} \qquad (1)$$

$$L_{NLoS}^{m,u} = L_{FS}^{m,u} + \eta_{NLoS} \qquad (2)$$

where $L_{FS}^{m,u}$ represents the free space pathloss which can be expressed as $L_{FS}^{m,u} = 20\log(4\pi f_c d_{m,u}/c)$, where f_c is the carrier frequency, c is the speed of light, and $d_{m,u}$ is the distance between mAeBS m and ground user u, which can be measured by coordinates. η_{LoS} and η_{NLoS} refers to the mean value of the excessive pathloss in LoS and NLoS links respectively.

The probability of LoS is a continuous function determined by the environment and elevation angle $\theta_{m,u}$ between mAeBS m and UE u:

$$P_{LoS}^{m,u} = \frac{1}{1 + a\exp\left(-b\left(\theta_{m,u} - a\right)\right)} \qquad (3)$$

where a and b are environment-related S-curve parameters. The elevation angle $\theta_{m,u}$ can be obtained as $\theta_{m,u} = \sin^{-1}(h_m/d_{m,u})$. The probability of NLoS can be obtained as $P_{NLoS}^{m,u} = 1 - P_{LoS}^{m,u}$.

Hence, the pathloss between mAeBS m and UE u can be obtained as:

$$L_{m,u} = \mathbb{I}\left(P_{LoS}^{m,u}\right) L_{LoS}^{m,u} + \left(1 - \mathbb{I}\left(P_{LoS}^{m,u}\right)\right) L_{NLoS}^{m,u} \qquad (4)$$

where $\mathbb{I}(r)$ is a Bernoulli random variable whose value is 1 with the probability r.

2. Beam scheduling

Besides the A2G propagation path loss, the directional mmWave antenna gain is also an important factor in the mAeBS channel. In this work, we suppose that a 3D beam has the same gain G_M within its beamwidth and a small constant sidelobe gain G_S outside the beamwidth, as shown in Fig.2 (a). The main lobe gain G_M can be obtained as follows [18]:

$$G_M = \frac{2 - \left(2 - \left(1 - \cos\delta_u\right)\right) G_S}{1 - \cos\delta_u} \qquad (5)$$

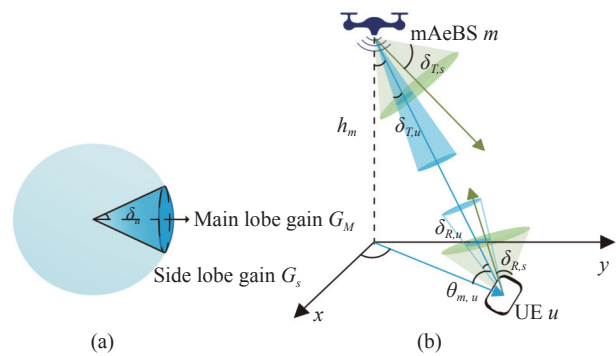where δ_u is the cone half angle of the mmWave beam for UE u.

Fig. 2. mAeBS serving UE with directional beam. (a) 3D Beam; (b) Beam alignment.

The main beams of the transmitter and receiver should be aligned before the data transmission starts, which will bring time overhead. We adopt the widely-used two-stage beam search scheme for beam alignment, in which the coarse-grained scanning is conducted on the sector level at first, and then the beam-level

scanning is carried out in the selected sector range. The beam alignment process is illustrated in Fig.2(b), where $\delta_{T,s}$ and $\delta_{R,s}$ depict the sector width at transmitter and receiver respectively, and $\delta_{T,u}$ and $\delta_{R,u}$ represent the beamwidth of the beam link from mAeBS m to UE u at both ends. The time cost of beam alignment is basically in the second stage as the time spent in the first stage is negligible. The beam alignment time τ is derived as [18]:

$$\tau = \frac{1 - \cos \delta_{T,s}}{1 - \cos \delta_{T,u}} \cdot \frac{1 - \cos \delta_{R,s}}{1 - \cos \delta_{R,u}} T_p \tag{6}$$

where T_p is the time when the beam traverses through the entire sector and sends a pilot signal at each position for alignment.

Due to the strong directivity of the narrow mm-Wave beam, it is is required to conduct beam scanning to cover the whole considered area. Assuming that each mAeBS casts N_b narrow beams in 3D space, when the number of associated UEs N_u is higher than the beams of the mAeBS, a round-robin scheme needs to be adopted for beam scheduling. The average ratio of time-frequency resources η_u occupied by UE u is approximated as

$$\eta_u = \begin{cases} 1, & N_b \leq N_u \\ \dfrac{N_b}{N_u}, & N_b > N_u \end{cases} \tag{7}$$

3. Capacity evaluation model under cooperative scheme

Unlike fixed ground BSs, mAeBSs have a fluid arrangement, so their coverage may often overlap and interference from other mAeBSs will occur. Therefore, an interference reduction method is required. To address this issue, we introduce CoMP technology to suppress the significant interference. We assume that multiple mAeBSs provide services for ground UEs collaborately, forming n cooperative mAeBS-UE clusters. In ith cooperative cluster, the set of mAeBSs and UEs are denoted as $\boldsymbol{M}_i$ and $\boldsymbol{U}_i$, respectively, and $\boldsymbol{M}_1 \cup \cdots \boldsymbol{M}_i \cdots \cup \cdots \boldsymbol{M}_n = \boldsymbol{M}$, $\boldsymbol{U}_1 \cup \cdots \boldsymbol{U}_i \cdots \cup \cdots \boldsymbol{U}_n = \boldsymbol{U}$.

We divide the mAeBSs with strong interference into a cooperative set and the joint transmission CoMP (JT-CoMP) scheme is adopted within the cooperative BS cluster. With JT-CoMP, from the point of view of the selected UE, the signal from other mAeBSs is not treated as interference but the desired signal [8]. Therefore, the strong interference is eliminated and the useful signal is increased, and the signal to interference and noise ratio (SINR) ξ_u of the signal received at UE u in the ith cooperative set can be approximated as:

$$\xi_u = \frac{\displaystyle\sum_{m \in \boldsymbol{M}_i} P_m G_{T,M} G_{R,M} L_{m,u}^{-1}}{\sigma^2}, u \in \boldsymbol{U}_i \tag{8}$$

where P_m is the transmit power of mAeBS m, σ^2 denotes the power of thermal noise, $G_{T,M}$ and $G_{R,M}$ are the main lobe antenna gain of transmitter and receiver.

Then the throughput of user u can be expressed as

$$\varrho_u = \eta_u \left(1 - \frac{\tau}{T} \right) B \log_2 \left(1 + \xi_u \right) \tag{9}$$

where B represents the channel bandwidth, τ is the beam alignment time in (6), T is the time slot and τ needs to be less than one slot T to ensure enough time for data transmission.

The system throughput can be expressed as

$$\varrho = \sum_{i=1}^{n} \left[\sum_{u \in \boldsymbol{U}_i} \varrho_u \right] \tag{10}$$

4. Problem formulation

In this paper, we aim to adjust the 3D placement of mAeBSs and the cooperative mAeBS-UE associations to achieve optimal system throughput. The optimization problem is formulated as follows:

$$\max_{x_m, y_m, h_m, n, \boldsymbol{M}_n, \boldsymbol{U}_n} \varrho$$

$$\begin{aligned} \text{s.t.} \quad & \text{C1: } x_{\min} < x_m < x_{\max}, \forall m \in \boldsymbol{M} \\ & \text{C2: } y_{\min} < y_m < y_{\max}, \forall m \in \boldsymbol{M} \\ & \text{C3: } h_{\min} < h_m < h_{\max}, \forall m \in \boldsymbol{M} \\ & \text{C4: } (x_m, y_m, h_m) \neq (x_l, y_l, h_l), \forall m, l \in \boldsymbol{M}, m \neq l \\ & \text{C5: } \boldsymbol{M}_1 \cup \cdots \cup \boldsymbol{M}_n = \boldsymbol{M} \\ & \text{C6: } \boldsymbol{U}_1 \cup \cdots \cup \boldsymbol{U}_n = \boldsymbol{U} \end{aligned}$$

$$(11)$$

where ϱ is the system throughput in (10). C1–C3 constraint mAeBSs from flying out of the area, and C4 is the collision constraint that prevents mAeBSs from flying to the same position. C5 and C6 ensure that all users in the system can be covered.

IV. Two-Timescale Framework for User Association and 3D Deployment of Cooperative mAeBSs

In this section, a two-timescale approach, as shown in Fig.3, is designed to solve the problem in (11). The whole process can be divided into two procedures in different timescales. On the large timescale, the mAeBSs periodically perform APC based on the large-scale interference characteristics, and the cooperative mAeBS-UE association is established. On the small timescale, a MADDPG-based deployment algorithm is designed to move mAeBSs to achieve optimal throughput. Specific-

ally, each mAeBS selects its actions based on its own observation in each subframe, aiming to increase the reward, which is related to the large timescale mAeBS-UE association.
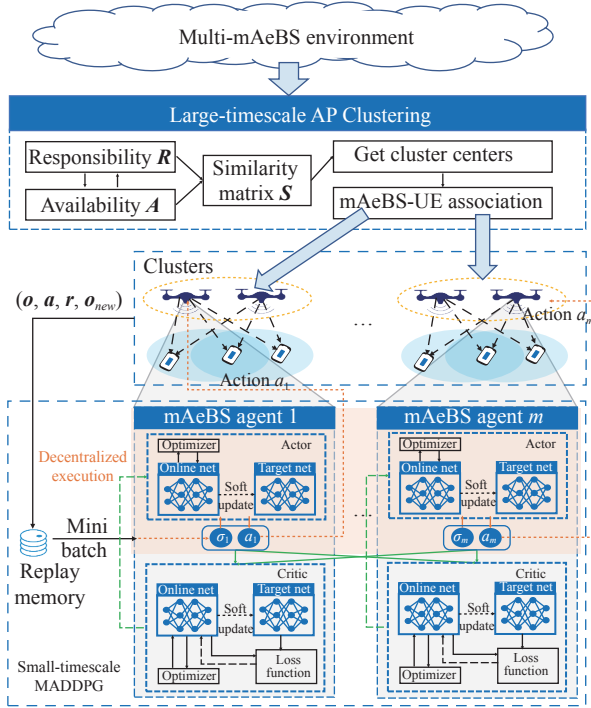

Fig. 3. The framework of the proposed approach.

1. APC-based cooperation mechanism for mAeBSs

We adopt APC for mAeBS-UE partitioning, considering the interference characteristics. APC clusters data points based on a similarity matrix $\boldsymbol{S}$. Generally, APC uses negative Euclidean distance to represent similarity, so the higher the similarity, the closer the distance between two points. In this paper, the similarity is calculated to characterize the interference between mAeBSs, so that the strong interfering mAeBSs can be clustered into a cluster. After APC, mAeBSs within a cluster serve their UEs collaboratively, transforming strong interference signals into useful signals.

Considered a network of M mAeBSs, the similarity matrix is denoted as

$$\boldsymbol{S} = [s_{m,l}]_{M \times M} \tag{12}$$

where $s_{m,l}$ indicates the interference of mAeBS l to mAeBS m associated users, $s_{m,l} = \sum_{U_m} P_m G_{T,S} G_{R,S} L_{m,i}^{-1}$, where $G_{T,S}$ and $G_{R,S}$ are the side lobe antenna gain of transmitter and receiver.

Instead of specifying the number of clusters in advance, APC treats all mAeBSs as potential cluster centers, called *exemplars*. The core idea of APC is to select the exemplar by continuously passing responsibil-

ity $\boldsymbol{R} = [r_{m,l}]_{M \times M}$ and availability $\boldsymbol{A} = [a_{m,l}]_{M \times M}$, between data points. $r_{m,l}$ indicates whether mAeBS l is suitable to be the exemplar of mAeBS m, and $a_{m,l}$ reflects the suitability of mAeBS m to choose mAeBS l as its exemplar. The stronger $r_{m,l}$ and $a_{m,l}$ are, the higher the probability that mAeBS l is the exemplar, and the higher the probability that mAeBS m belongs to the cluster with mAeBS l as the exemplar. $r_{m,l}$ and $a_{m,l}$ are iteratively updated and the update equations are expressed as [27]:

$$r_{m,l}^{t+1} = s_{m,l} - \max_{l' \neq l} \left\{ a_{m,l'}^t + s_{m,l'} \right\} \tag{13}$$

$$a_{m,l}^{t+1} = \min \left(r_{l,l}^{t+1} + \sum_{m' \notin \{m,l\}} \max \left\{ 0, r_{m',l}^{t+1} \right\}, 0 \right) \tag{14}$$

The detailed procedures of our proposed APC-based cooperative mAeBS-UE association is illustrated in Algorithm 1.

First of all, each UE associates with the mAeBS with the strongest received signal in the initial stage. At this time, each mAeBS serves a set of UEs, we refer it to the initial UE set $\boldsymbol{N}_u$. Then, the modified APC is conducted. When the algorithm converges, mAeBSs will be divided into several clusters according to the interference characteristics. mAeBSs within each cluster perform cooperative communication, and the associated UE set $\boldsymbol{U}_n$ of each mAeBS is a collection of the initial UE set $\boldsymbol{N}_u$ of the mAeBSs within the cluster.

Algorithm 1   APC for mAeBSs cooperation

Input: Positions of mAeBSs and UEs.

Output: mAeBS-UE cooperative association $\boldsymbol{M}_n$ and $\boldsymbol{U}_n$.

Initialize: Responsibility $\boldsymbol{R} = [0]_{M \times M}$ and availability $\boldsymbol{A} = [0]_{M \times M}$

1: Get initial mAeBS-UE association: UE u chooses the mAeBS with the strongest receiving power, namely, $\text{argmax}_{m \in \boldsymbol{M}} P_m G_{T,M} G_{R,M} L_{m,u}^{-1}$. Each mAeBS serves a set of UEs $\boldsymbol{N}_u$.

2: Calculate similarity $\boldsymbol{S}$ according to (12).

3: REPEAT

4: Update $\boldsymbol{R} = [r_{m,l}]$ and $\boldsymbol{A} = [a_{m,l}]$ according to (13) and (14) and broadcast.

5: UNTIL max iteration or the APC algorithm converges

6: FOR each mAeBS $m \in \boldsymbol{M}$

7: Select the mAeBS l with the largest sum of responsibility and availability as its exemplar, namely, $\text{exemplar}(m) = \text{argmax}_{l \in \boldsymbol{M}} \{ a_{m,l} + r_{m,l} \}$.

8: Obtain clustered mAeBSs set $\boldsymbol{M}_n$, where n is the number of exemplars.

9: Obtain the associated UE set $\boldsymbol{U}_n$ according to $\mathcal{M}_n$ and initial UE set $\boldsymbol{N}_u$.

---

2. MADDPG for 3D Deployment of the mAeBSs

We model the proposed multi-mAeBS cooperation problem as a Markov game, which can be regarded as a multi-agent extension of the Markov decision process. In our proposed multi-mAeBS cooperation scheme, the basic components are defined as follows:

Agent Each mAeBS can be viewed as an agent and each agent gets its own observation and selects actions according to its own policy, then receives a reward from the environment and moves to its next state.

Observation The observation of each mAeBS is the current 3D location of each mAeBS $o_m = \{(x_m, y_m, h_m)\}$. $\boldsymbol{o} = \{o_m, m \in \boldsymbol{M}\}$.

Action The action of each agent is the moving distance torward 6 directions $a_m = \{(\Delta x_m, \Delta y_m, \Delta h_m)\}$, i.e., left or right, forward or backward, and up or down. $\boldsymbol{a} = \{a_m, m \in \boldsymbol{M}\}$.

Reward In our proposed system, the reward is the system throughput calculated according to (10) after each mAeBS's location changes at time t. Since there is a cooperative relationship between mAeBSs, mAeBSs share the same reward at the same time.

MADDPG is an extended algorithm of deep deterministic policy gradient algorithm (DDPG) in multi-agent environments. DDPG is essentially realized by combining the ideas of DQN and Actor-Critic algorithms. In MADDPG, each agent has two modules: actor and critic. The actor is trained for generating a deterministic policy which chooses a random action from a determined distribution, and the critic is trained to simulate the real Q-table using neural networks. Both the actor and critic have an online network and a target network, and the target network has the same architecture as the online network and is introduced to solve the instability problem.

The actor gets the selection probability of the action according to the current state. In the game with M agents, the deterministic policies for all agents $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_M\}$ are parameterized by $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_M\}$. The actor network is updated by minimizing the gradient of the expected return for agent m, which can be written as:

$$\nabla_{\theta_m} J = \mathbb{E}\left[\nabla_{\theta_m} \mu_m\left(o_m^j\right) \right. \\ \left. \times \nabla_{a_m} Q_m^{\boldsymbol{\mu}}\left(\boldsymbol{o}^j, a_1^j, \ldots, a_M^j\right)\Big|_{a_m = \mu_m\left(o_m^j\right)}\right] \quad (15)$$

where μ_m represents the policy with respect to θ_m.

The critic approximates the value function of the state-action pair to judge the quality of the actor's selected action. It is updated by minimizing the following loss function:

$$L\left(\omega_m\right) = \mathbb{E}\left[y^j - Q_m^{\boldsymbol{\mu}}\left(\boldsymbol{o}^j, a_1^j, \ldots, a_M^j\right)\right]^2 \quad (16)$$

where y^j is the target value and can be estimated as:

$$y^j = r_m^j + \gamma \cdot Q_m^{\boldsymbol{\mu}'}\left(\boldsymbol{o}_{new}^j, a_1', \ldots, a_M'\right)\Big|_{a_m' = \mu_m'\left(o_m^j\right)} \quad (17)$$

where $\boldsymbol{\mu}'$ is the target policy set with parameter θ_m' and γ denotes the discount factor.

The complete process of our proposed MADDPG-based multi-mAeBS 3D deployment algorithm are shown in Algorithm 2. At first, each mAeBS initializes its own four deep neural networks with random weights, i.e., online actor network, target actor network, online critic network, and target critic network. The replay memory buffer which enables agents to remember and reuse past experiences is also initialized. Then, the exploration process is conducted in the training process. The action derived from the current actor network is added with noise N for exploration. The capacity under the large timescale mAeBS-UE association is employed to update the reward function. The transitions $(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{r}, \boldsymbol{o}_{new})$ in $\mathcal{D}$ are then stored in the replay memory and critic networks can thus obtain the observations and actions of all agents. By exploration, each mAeBS will choose an action with the highest reward and fly along this direction. Next, we use the mini-batch method to randomly collect samples from the replay memory and update the weight of the actor and critic network accordingly. In order to improve the stability of learning, the parameters of the target network are softly updated.

Algorithm 2 MADDPG-based Deployment of mAeBSs

Input: Locations of UEs; structures of actor and critic networks; training episodes; max steps of each episode.

Output: Trained actor network of each mAeBS.

Initialize: Online actor network μ_m and online critic network Q_m for each mAeBS m with random weights θ_m and ω_m, respectively; target actor network μ_m' and target critic network Q_m' for each mAeBS m with weights $\theta_m' = \theta_m$ and $\omega_m' = \omega_m$, respectively; replay memory $\mathcal{D}$ with capacity K.

1: For each training episode:

2:     Update the location of UEs if there is any change;

3:     Initialize the locations of mAeBSs;

4:     Initialize a random noise N for action exploration;

5:       For each step:

6:       Each mAeBS m obtains its observation o_m, and selects the action based on its own observation $a_m = \mu(o_m|\theta_m) + N$;

7:       Each mAeBS m sets its own location based on the action a_m and obtains new observation $\boldsymbol{o}_{new}$. Note that the mAeBS will stay at its current location if it

flies out of the region or it collides with another mAeBS;

8:     Every ϵ steps execute mAeBSs clustering based on $\boldsymbol{o}_{new}$ according to Algorithm 1;

9:     Observe reward $\boldsymbol{r}$;

10:    Store transitions $(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{r}, \boldsymbol{o}_{new})$ in $\mathcal{D}$;

11:    $\boldsymbol{o} \leftarrow \boldsymbol{o}_{new}$;

12:    For each mAeBS $m \in \boldsymbol{M}$:

13:      Sample a random minibatch of transitions $\left(\boldsymbol{o}^j, \boldsymbol{a}^j, \boldsymbol{r}^j, \boldsymbol{o}^j_{new}\right)$ from $\mathcal{D}$;

14:      Update the critic according to (16);

15:      Update the actor according to (15);

16:    Every C steps update target network for each mAeBS m: $\omega'_m \leftarrow \tau\omega_m + (1-\tau)\omega'_m$, $\theta'_m \leftarrow \tau\theta_m + (1-\tau)\theta'_m$.

At the execution stage, each mAeBS downloads its well-trained actor network and receives initial observation. Then, each mAeBS selects its action based on its own observation, without the need of communicating with other mAeBSs. The centralized training and distributed execution framework is especially suitable for UAV-assisted networks since it brings less communication burden to resource-limited UAVs. The time-consuming centralized training process can be conducted at the aerial base station controller (ABC). Each mAeBS downloads the well-trained actor network from the ABC and makes motion decisions based on its own observation. The proposed autonomous deployment management architecture of mAeBSs is shown in Fig.4.
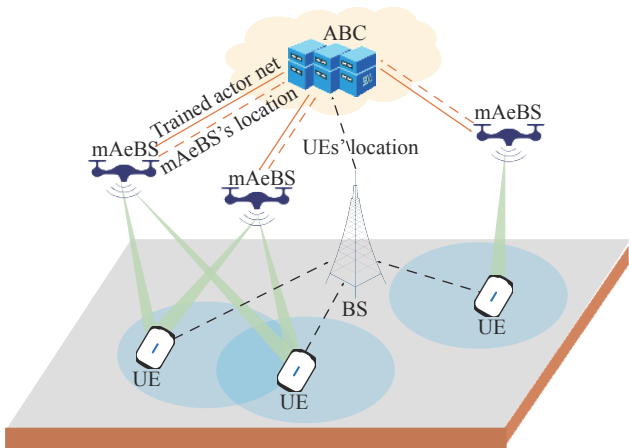

Fig. 4. The architecture of autonomous deployment management of mAeBSs.

The complexity of the proposed mechanism is mainly composed of two parts: MADDPG algorithm complexity and APC algorithm complexity. The time complexity of the APC algorithm is $O(M^2 I)$, where M is the number of mAeBSs, and I is the iteration times. The time complexity of MADDPG is determined by the operations in deep neural networks. We assume that an actor network has N_h^a hidden layers, which have q_i^a

neurons respectively, $i = 1, \ldots, N_h^a$. Similarly, a critic network has N_h^c hidden layers, and the jth hidden layer contains q_j^c neurons. And in each iteration, the time complexity of MADDPG can be obtained as $O\left(\sum_{i=0}^{N_h^a-1} q_i^a q_{i+1}^a + \sum_{j=0}^{N_h^c-1} q_j^c q_{j+1}^c\right)$ [36]. Here, the operations at input and output layers are ignored since they are relatively trivial. In the execution process, the time complexity can be reduced to $O\left(\sum_{i=0}^{N_h^a-1} q_i^a q_{i+1}^a\right)$ since no critic network is needed.

V. Simulation Results and Discussions

In this section, the performance of our proposed scheme is evaluated. We consider a square area of 6 km $\times$ 6 km, in which ground UEs follow the 2D Gaussian distribution to model the user aggregation in the real hotspot scenario. The vertical flying region of mAeBSs is [10 m, 200 m]. The simulations are implemented based on the python 3.8 and torch 1.7.0+cu110 environment. The main parameters of the environment and hyperparameters of MADDPG are listed in Tables 2 and 3, respectively.

Table 2. Main parameters of the environment

Parameters	Values
Carrier frequency	30 GHz
Channel bandwidth B	100 MHz
Transmit power of mAeBSs P_m	50 dBm
Side lobe gain $G_{T,S}, G_{R,S}$	-10 dB, -10 dB
Sector width $\delta_{T,s}, \delta_{R,s}$	$\pi/2$, $\pi/2$
Beamwidth $\delta_{T,u}, \delta_{R,u}$	$\pi/6$, $\pi/6$
S-curve parameters a, b	9.61, 0.16
Excessive pathloss η_{LoS}, η_{NLoS}	1, 20
Thermal noise power density σ^2	-174 dBm/Hz
Pilot duration ratio T_p/T	2×10^{-4}

Table 3. Hyperparameters of the MADDPG

Hyperparameters	Values
Layers of actor, critic	4, 4
Learning rate of actor, critic	0.01, 0.01
Layer type of actor	Fully connected
Neurons of hidden layers for actor	[64,64]
Layer type of critic	Fully connected
Neurons of hidden layers for critic	[64,64]
Optimizer	Adam
Activation function	Leaky ReLU
Memory capacity K	10000
Minibatch size	1256
Discount factor γ	0.97
Soft updating rate	0.5

First of all, we compare our proposed MADDPG with two baseline DRL algorithms, i.e., DQN and multi-agent DQN (MADQN). DQN is a widely-adop-

ted single-agent DRL algorithm. To make the DQN suitable for the proposed scenario, we discretize the action space and define it as moving the mth mAeBS forward, backward, left, right, up, or down, a total of 6M options. MADQN is an extended version of DQN in the multi-agent environment. Fig.5 shows the averaged rewards obtained by these 3 algorithms in the training and execution process. The training process is conducted in the first 400 episodes. In the next 400 episodes, the trained networks are uploaded to mAeBS agents and agents make action decisions accordingly. Results show that rewards of three algorithms increase gradually and MADDPG can achieve a higher reward since its continuous action space makes it possible to obtain more accurate results. In the execution process, all the three algorithms can stay at a stable and high level. However, different from DQN and MADQN, agents in MADDPG do not need global observations in the execution process. It means that mAeBS does not need to communicate with each other when deciding its action after being fully trained, which is practical in the real multi-mAeBS network.
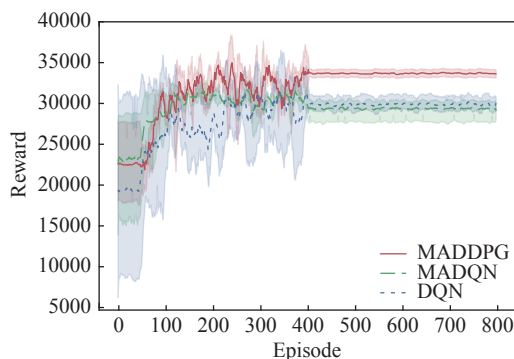

Fig. 5. Rewards versus episodes in training and execution process.

We also compare MADDPG with DQN and MADQN in training time, as shown in Fig.6. As the result shows, multi-agent DRL consumes less time to train than centralized single-agent DRL. In terms of running time, compared with multi-agent algorithms, DQN needs more training time due to its large action space, and this gap has the tendency to be wider with the increase of the number of mAeBSs.

Fig.7 shows the impact of the number of mAeBSs on the capacity performance. We compare our proposed scheme with two baseline scenarios: DQN with APC and MADDPG without APC. In the DQN with APC scheme, mAeBS-UE cooperative association is obtained through APC, similarly to the proposed scheme, and a central controller makes decisions through DQN to move mAeBSs for better throughput. In the MADDPG without APC scheme, each UE associates with the mAeBS which is with the strongest signal, i.e.,
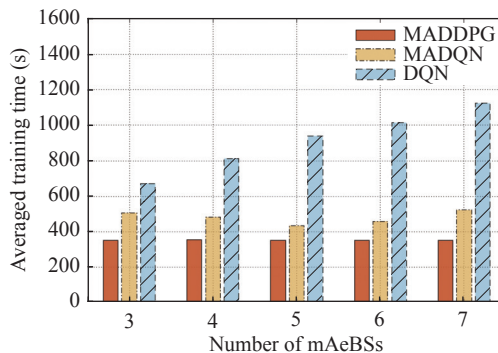

Fig. 6. Computational time of different algorithms under different numbers of mAeBSs.

CoMP is not introduced in this scheme, and mAeBSs make decisions through MADDPG to move to the place where the system throughput can be maximized. Results show that in both the proposed and DQN with APC scheme, the throughput per UE is generally on the rise with the increase of the number of mAeBSs, while the increase of mAeBSs has a negative impact on the throughput of each user in the MADDPG without APC scheme. This is because more mAeBSs bring more interference to the network and thus deteriorate the capacity performance, and the introduction of CoMP can solve this problem. In addition, compared to DQN with APC, the proposed scheme gains a better performance with the help of MADDPG.
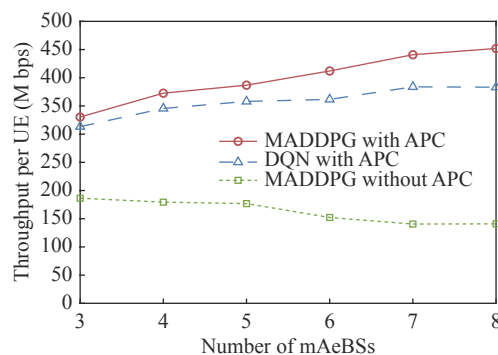

Fig. 7. Averaged throughput for different numbers of mAeBSs.

Apart from the capacity performance, we also analyze the coverage performance, as shown in Fig.8. We utilize coverage probability to characterize coverage performance, which is defined as the probability that the SINR received by the UE is greater than a given threshold. As result shows, when the SINR threshold is set as 10 dB, only 10% UEs can be covered when three mAeBS are deployed under the MADDPG without APC scheme and the situation is even worse as more mAeBSs deployed. When CoMP is introduced, the covered UE increases to 50% under the same condition. In addition, compared with DQN, the application of MADDPG can also improve the system performance.
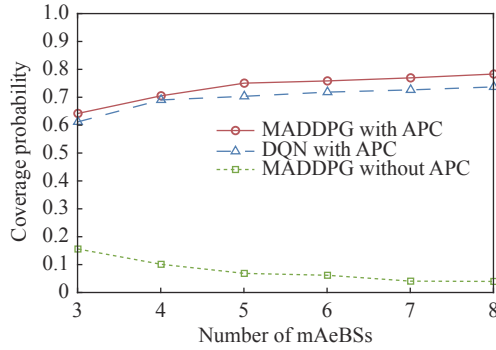
Fig. 8. Coverage probability with different numbers of mAeBSs.

In order to verify the capacity enhancement effect of CoMP in this scenario, the proposed scheme is compared with the MADDPG without APC scheme. Fig.9 shows the effect of the number of beams of mAeBSs and the density of UEs on the throughput performance. As the density of UEs increases, the throughput of each UE of the two schemes gradually declines. In the case of a small number of users, the performance of the proposed scheme is 130%–190% better than that of the MADDPG without APC scheme, but with the increase of the user density, the gap is gradually narrowing. In addition, compared with the MADDPG without APC scheme, the number of beams has a greater impact on the proposed scheme, since each mAeBSs associates with more users in the cooperative mechanism, so more beams are needed to provide services for them. Therefore, although the proposed scheme can significantly increase the throughput performance, it is at the cost of the beam resources of mAeBSs.
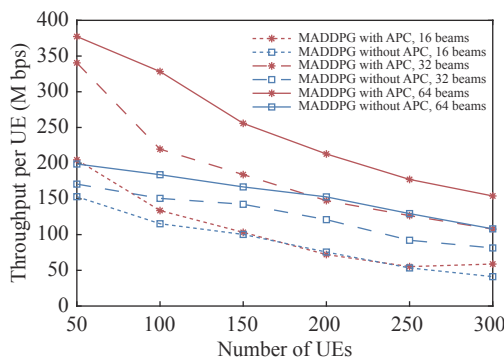

Fig. 9. Averaged throughput under different numbers of UEs.

Fig.10 shows the relationship between coverage probability and SINR thresholds, which is helpful to observe the bearing capacity of mAeBS systems for different scenarios, since different services may have different communication threshold requirements. As results show, the coverage probability of the proposed cooperative scheme is higher under the same SINR threshold and remains stable over a wider range compared to the MADDPG without APC scheme. Additionally, increas-

ing the number of deployed mAeBS is helpful to improve the coverage performance in our proposed scheme while it will deteriorate the performance of the MADDPG without APC scheme since more interference is introduced.
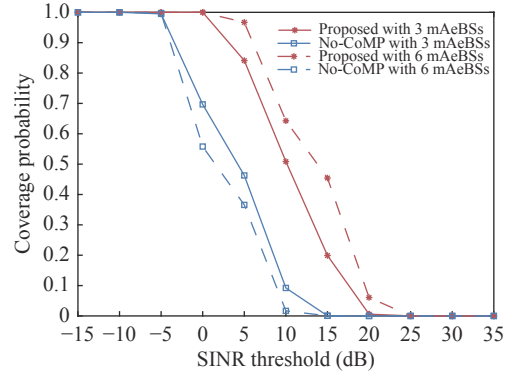

Fig. 10. Coverage probability under different SINR thresholds.

It can be seen from the above analysis that the proposed cooperation scheme can improve the capacity and coverage performance of multi-mAeBS networks, and the proposed MADDPG-based deployment algorithm has the advantages of less training time and higher converged reward in the multi-mAeBS scenario.

VI. Conclusions

This paper studies the deployment and UE association problem of using cooperative mAeBSs to enhance capacity for hotspot areas. We propose a two-timescale mechanism for capacity maximization in the multi-mAeBS network, in which APC is adopted to obtain the cooperative mAeBS-UE association according to the interference characteristics and MADDPG is used to make deployment decisions of mAeBSs. In our proposed scheme, after being fully trained, each mAeBS can make appropriate decisions based on its own observations, without the need of communicating with each other. Simulation results show the superiority of our proposed MADDPG-based multi-mAeBS deployment algorithm in convergency and time efficiency by comparing with baseline algorithms. Results also demonstrate that the proposed cooperative scheme can greatly improve the coverage and capacity of the network since strong interference is eliminated.

References

[1] L. Zhu, Z. Xiao, X. -G. Xia, *et al.*, "Millimeter-wave communications with non-orthogonal multiple access for B5G/6G," *IEEE Access*, vol.7, pp.116123–116132, 2019.

[2] Z. Xiao, L. Zhu, Y. Liu, *et al.*, "A survey on millimeter-wave beamforming enabled UAV communications and networking," *IEEE Communications Surveys & Tutorials*, vol.24, no.1, pp.557–610, 2022.

[3] Z. Xiao, L. Zhu, and X. -G. Xia, "UAV communications with millimeter-wave beamforming: Potentials, scenarios, and challenges," *China Communications*, vol.17, no.9, pp.147–166, 2020.

[4] L. Zhang, H. Zhao, S. Hou, *et al.*, "A survey on 5G millimeter wave communications for UAV-assisted wireless networks," *IEEE Access*, vol.7, pp.117460–117504, 2019.

[5] X. Zhao, Y. Zhang, P. Qin, *et al.*, "Key technologies and development trends for a space-air-ground integrated wireless optical communication network," *Acta Electronica Sinica*, vol.50, no.1, pp.1–17, 2022. (in Chinese)

[6] L. Zhu, J. Zhang, Z. Xiao, *et al.*, "Optimization of multi-UAV-BS aided millimeter-wave massive MIMO networks," in *Proceedings of 2020 IEEE Global Communications Conference*, Taipei, China, pp.1–6, 2020.

[7] S. Kumar, S. Suman, and S. De, "Dynamic resource allocation in UAV-enabled mmWave communication networks," *IEEE Internet of Things Journal*, vol.8, no.12, pp.9920–9933, 2021.

[8] D. Lee, H. Seo, B. Clerckx, *et al.*, "Coordinated multipoint transmission and reception in LTE-advanced: Deployment scenarios and operational challenges," *IEEE Communications Magazine*, vol.50, no.2, pp.148–155, 2012.

[9] Q. Cui, H. Song, H. Wang, *et al.*, "Capacity analysis of joint transmission CoMP with adaptive modulation," *IEEE Transactions on Vehicular Technology*, vol.66, no.2, pp.1876–1881, 2017.

[10] P. Peng, F. Zhu, Q. Liu, *et al.*, "Achieving safe deep reinforcement learning via environment comprehension mechanism," *Chinese Journal of Electronics*, vol.30, no.6, pp.1049–1058, 2021.

[11] H. Bayerlein, R. Gangula, and D. Gesbert, "Learning to rest: A Q-learning approach to flying base station trajectory design with landing spots," in *Proceedings of 2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp.724–728, 2018.

[12] P. Yu, J. Guo, Y. Huo, *et al.*, "Three-dimensional aerial base station location for sudden traffic with deep reinforcement learning in 5G mmWave networks," *International Journal of Distributed Sensor Networks*, vol.16, no.5, DOI: 10.1177/1550147720926374, 2020.

[13] Q. Wang, W. Zhang, Y. Liu, *et al.*, "Multi-UAV dynamic wireless networking with deep reinforcement learning," *IEEE Communications Letters*, vol.23, no.12, pp.2243–2246, 2019.

[14] X. Liu, Y. Liu, Y. Chen, *et al.*, "Machine learning aided trajectory design and power control of multi-UAV," in *Proceedings of 2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, pp.1–6, 2019.

[15] Z. Xiao, P. Xia, and X. -G. Xia, "Enabling UAV cellular with millimeter-wave communication: potentials and approaches," *IEEE Communications Magazine*, vol.54, no.5, pp.66–73, 2016.

[16] S. G. Sanchez, S. Mohanti, D. Jaisinghani, *et al.*, "Millimeter-wave base stations in the sky: An experimental study of UAV-to-ground communications," *IEEE Transactions on Mobile Computing*, vol.21, no.2, pp.644–662, 2022.

[17] M. Gapeyenko, D. Moltchanov, S. Andreev, *et al.*, "Line-of-sight probability for mmWave-based UAV communications in 3D urban grid deployments," *IEEE Transactions on Wireless Communications*, vol.20, no.10, pp.6566–6579, 2021.

[18] L. Zhu, J. Zhang, Z. Xiao, *et al.*, "3D beamforming for flexible coverage in millimeter-wave UAV communications," *IEEE Wireless Communications Letters*, vol.8, no.3, pp.837–840, 2019.

[19] H. -L. Chiang, K. -C. Chen, W. Rave, *et al.*, "Machine-learning beam tracking and weight optimization for mm-Wave multi-UAV links," *IEEE Transactions on Wireless Communications*, vol.20, no.8, pp.5481–5494, 2021.

[20] F. Zhou, W. Li, L. Meng, *et al.*, "Capacity enhancement for hotspot area in 5G cellular networks using mmWave aerial base station," *IEEE Wireless Communications Letters*, vol.8, no.3, pp.677–680, 2019.

[21] X. Guo, Y. Chen, and Y. Wang, "Learning-based robust and secure transmission for reconfigurable intelligent surface aided millimeter wave UAV communications," *IEEE Wireless Communications Letters*, vol.10, no.8, pp.1795–1799, 2021.

[22] L. Zhu, J. Zhang, Z. Xiao, *et al.*, "Millimeter-wave NOMA with user grouping, power allocation and hybrid beamforming," *IEEE Transactions on Wireless Communications*, vol.18, no.11, pp.5065–5079, 2019.

[23] Y. Liu, K. Liu, J. Han, *et al.*, "Resource allocation and 3-D placement for UAV-enabled energy-efficient IoT communications," *IEEE Internet of Things Journal*, vol.8, no.3, pp.1322–1333, 2021.

[24] X. Liu, Y. Liu, and Y. Chen, "Reinforcement learning in multiple-UAV networks: Deployment and movement design," *IEEE Transactions on Vehicular Technology*, vol.68, no.8, pp.8036–8049, 2019.

[25] I. Valiulahi and C. Masouros, "Multi-UAV deployment for throughput maximization in the presence of co-channel interference," *IEEE Internet of Things Journal*, vol.8, no.5, pp.3605–3618, 2021.

[26] C. Shen, T. Chang, J. Gong, *et al.*, "Multi-UAV interference coordination via joint trajectory and power control," *IEEE Transactions on Signal Processing*, vol.68, pp.843–858, 2020.

[27] B J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol.315, no.5814, pp.972–976, 2007.

[28] H. Zhang, C. Jiang, J. Cheng, *et al.*, "Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks," *IEEE Wireless Communications*, vol.22, no.3, pp.92–99, 2015.

[29] H. Zhang, H. Liu, C. Jiang, *et al.*, "A practical semidynamic clustering scheme using affinity propagation in cooperative picocells," *IEEE Transactions on Vehicular Technology*, vol.64, no.9, pp.4372–4377, 2015.

[30] J. Zhang, G. Chuai, and W. Gao, "Power control and clustering-based interference management for UAV-assisted networks," *Sensors*, vol.20, no.14, article no.3864, 2020.

[31] C. Qiu, Z. Wei, X. Yuan, *et al.*, "Multiple UAV-mounted base station placement and user association with joint fronthaul and backhaul optimization," *IEEE Transactions on Communications*, vol.68, no.9, pp.5864–5877, 2020.

[32] A. Alzidaneen, A. Alsharoa, and M. Alouini, "Resource and placement optimization for multiple UAVs using backhaul tethered balloons," *IEEE Wireless Communications Letters*, vol.9, no.4, pp.543–547, 2020.

[33] R. Lowe, Y. Wu, A. Tamar, *et al.*, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, Long Beach, CA, USA, pp.6382–6393, 2017.

[34] Y. Zhang, Z. Mou, F. Gao, *et al.*, "UAV-enabled secure communications by multi-agent deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol.69, no.10, pp.11599–11611, 2020.

[35] L. Wang, K. Wang, C. Pan, *et al.*, "Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing," *IEEE Transactions on Cognitive Communications and Networking*, vol.7, no.1, pp.73–84, 2021.

[36] A. Gao, Q. Wang, W. Liang, *et al.*, "Game combined multi-

agent reinforcement learning approach for UAV assisted off-loading," *IEEE Transactions on Vehicular Technology*, vol.70, no.12, pp.12888–12901, 2021.

[37] J. Yao and J. Xu, "Joint 3D maneuver and power adaptation for secure UAV communication with CoMP reception," *IEEE Transactions on Wireless Communications*, vol.19, no.10, pp.6992–7006, 2020.

[38] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol.3, no.6, pp.569–572, 2014.

FENG Lei (corresponding author) received the B.E. and Ph.D. degrees in communication and information systems from BUPT in 2009 and 2015. He is an Associated Professor at present in State Key Laboratory of Networking and Switching Technology, BUPT. His research interests are resources management in wireless network and smart grid.

(Email: fenglei@bupt.edu.cn)

ZHAO Yikun received the B.E. degree in communication engineering from Inner Mongolia University in 2019. She is currently pursuing the Ph.D. degree in computer science and technology in Beijing University of Posts and Telecommunications (BUPT). Her research interest is in aerial base station networks. (Email: yizhao@bupt.edu.cn)

LI Wenjing is a Professor at BUPT and serves as a Director in the Key Laboratory of Network Management Research Center. Meanwhile, she is the Leader of TC7/WG1 in the China Communications Standards Association (CCSA). Her research interests are wireless network management and automatic healing in SONs.

ZHOU Fanqin received the Ph.D. degree in automation from BUPT, China, in 2019. He is currently a Lecturer with the State Key Laboratory of Networking and Switching Technology in BUPT. His current research interests include intelligent network slicing and resource management of mobile edge networks.

YU Peng received the B.E. and Ph.D. degrees from BUPT in 2008 and 2013 respectively. He is an Associate Professor at present in State Key Laboratory of Networking and Switching Technology, BUPT. His research interests are intelligent and green network management for 5G/6G networks and smart grid communication networks.