3D Deployment and User Association of CoMP-assisted Multiple Aerial Base Stations for Wireless Network Capacity Enhancement

Yikun Zhao, Fanqin Zhou, Wenjing Li*, Yuan Gao, Lei Feng*, Peng Yu
State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, China
*{wjli, fenglei}@bupt.edu.cn

Abstract—Deploying aerial base stations (AeBSs) has been regarded as an effective solution to wireless network capacity enhancement in specific areas with excessive traffic burden but insufficient capacity. Since the traffic distributions in wireless networks tend to be ever-changing, the deployed AeBS need to adjust its position to rapidly and continuously adapt to the drifting capacity enhancement demands, which is difficult to handle with traditional optimization methods due to high computational complexity and poor scalability. In this paper, we design a multi-agent deep reinforcement learning-based 3D AeBS deployment algorithm with the goal of maximizing the system throughput, which is able to make decisions in dynamic environments and conducts in a distributed manner. Additionally, in order to address the interference issue between multiple AeBSs, we adopt the Coordinated Multiple Points Transmission (CoMP) in the air-to-ground communication and propose a clustering algorithm to form groups of AeBSs for cooperative communication based on the network interference characteristics. Simulation results demonstrate that the proposed approach has significant throughput gains over conventional schemes without CoMP, and that the proposed multi-agent deep Q network (MADQN) is more efficient than centralized DQN in deriving the solution.

Index Terms—Unmanned aerial vehicle; cooperative communication; capacity enhancement; multi-agent deep reinforcement learning.

I. INTRODUCTION

With the development of various emerging applications in the era of 5G, users have higher requirements for mobile network bandwidth and transmission rates [1]. Especially when terrestrial base stations are destroyed or overloaded and can hardly meet the traffic demand of the user terminals, it is necessary to enhance the network capacity of the current area. With unique advantages of flexible mobility, ever-reducing cost and high possibilities of line-of-sight (LoS) links, Unmanned Aerial Vehicle (UAV) has been widely investigated to serve as aerial base station (AeBS) in wireless communication networks [2]. An aerial base station can freely adjust its location according to the distribution and traffic demand of users, and has been regarded as an effective supplement to the existing cellular system to improve the ground wireless capacity and coverage under the ultra intensive service demand.

Generally, a single AeBS can only provide limited coverage, while a cooperative network of multiple AeBSs can effectively expand the coverage area and increase the number of served users [3]. However, interference is a critical factor in multi-AeBS networks, which severely affects network communication performance. Coordinated Multiple Points Transmission (CoMP) has been widely used as an effective approach to control the interference in traditional cellular networks [4]. With CoMP technology, each user associates with several BSs collaboratively, transforming significant interference into cooperative signal sources and maintaining sufficiently reliable communication links. Rich works have shown that it can improve the transmission performance in terrestrial cellular networks [4]–[6]. However, existing work on CoMP for supporting AeBS-integrated networks is still limited.

Optimal 3D placement of multiple coordinating AeBSs is also a challenging issue. Many works utilize traditional optimization algorithms to solve this problem [7]–[9], which have the problem of high computational complexity and can not guarantee that the computed policy is still the optimal policy once the network changes. Considering the dynamic characteristic in multi-AeBS network, these algorithms are difficult to be applied in practical scenarios. Deep Reinforcement Learning (DRL), a most trending type of Machine Learning (ML), enables a paradigm for decision systems to accumulate and utilize experience in changing environments. Many works have investigated problems in AeBS-assisted wireless networks by invoking DRL algorithms [10]–[14]. However, these studies conduct the proposed DRL algorithms in a centralized way, i.e., assumes that there is a central controller which can obtain all information and decide which UAV take what action. These centralized DRL approaches are inconvenient due to their large state space and action space. Fortunately, these challenges can be addressed by applying Multi Agent Reinforcement Learning (MARL) algorithms, which handle difficulties in modelling and computation in distributed manners [15].

In this work, we propose a 3D deployment and user association scheme for CoMP-assisted multi-AeBS communication. The main contribution of this work can be organized as follows. First of all, a capacity evaluation model of multi-AeBS in the cooperative communication is established. Next, in order to address the network interference characteristics,

an Affinity Propagation (AP) clustering-based AeBS-UE partitioning algorithm is introduced to divide AeBSs and UEs into different cooperative clusters. With this algorithm, AeBSs severely interfering with each other are divided into a cluster and provide service for their associated ground users collaboratively, converting strong interference into collaborative signals, thus improving system capacity performance. Finally, a multi-agent DQN (MADQN) algorithm, which conducts in a distributed manner and is more suitable for multi-AeBS environments, is designed for the 3D deployment of multi-AeBS with the goal of maximizing the system throughput.

The rest of the paper is organized as follows. Section II presents the related works. In Section III, the system model of our studied cooperative multi-AeBS is introduced. Then, a clustering-based cooperation mechanism and a MADQN-based deployment scheme are designed to maximize the system capacity in Section IV. Next, simulation results are presented and discussed in Section V. Finally, Section VI concludes this paper.

II. Related Work

Early research on AeBS-assisted networks mainly focuses on the case of a single AeBS deployed. Recently, as the distribution of users becomes denser and the demands for communication quality of users increase rapidly, deploying a single AeBS can hardly meet the communication requirements. Therefore, lots of attention has been paid to the multi-AeBS network. Literature [12] and [16] obtain the cell partition of the users using K-means algorithm and a genetic algorithm based K-means respectively, and set cluster centers as the initial locations of the AeBSs. The above works divide AeBSs and users into different clusters based on geographical information, ignoring the potential problem brought by interference to simplify the model. However, interference is an inevitable problem in multi-AeBS networks. In [17], the authors aim to maximize the minimum achievable system throughput considering co-channel interference. Reference [18] studies the cross-link interference between AeBSs and their associated ground users and jointly optimizes trajectory and power control to achieve a higher system sum rate. Although these works address the problem of interference in multi-AeBS networks, they do not consider the utilization of the CoMP technique to transform the interference to capacity gain.

The 3D deployment design of UAVs is a significant and fundamental topic. Recently, much research has designed algorithms for the collaborative deployment of multiple AeBSs. In [7], a polynomial-time successive AeBS placement scheme is proposed, aiming to minimize the number of AeBSs while ensuring that each user is covered by at least one AeBS. The authors in [8] develope an iterative algorithm to jointly optimize AeBSs' placement and user association with the aid of the gradient ascent, dual-domain coordinated descent, and bipartite graph matching. Literature [9] aims to maximize the throughput by optimizing the deployment, transmit power, and associations of the AeBSs considering backhauling constraints,

and proposes an efficient heuristic algorithm based on shrink-and-realign process to optimize the placement of AeBSs. However, the above algorithms require global environment information to calculate the optimal solution, and the results need to be recalculated once the network topology changes. Due to the mobility of AeBSs, their network topology environment tends to change rapidly, and the process of recalculating the optimal results is too slow for real-time operation, which greatly limits the application of these algorithms in real multi-AeBS environments.

To handle these problems, many works have investigated the deployment design problem of UAVs by invoking DRL algorithms. Deep Q Network (DQN), a popular DRL algorithm, has been widely used in the deployment problem of UAVs [11]. In [11], a UAV positioning scheme is designed based on DQN to determine the optimal link between two UAVs. In order to achieve better performance, some works adopt algorithm variations of DQN for the flight control of UAVs, such as Dueling DQN and Prioritized Replay Double DQN. Literature [13] utilizes DDQN to optimize the deployment of multiple AeBSs and results validate that the proposed algorithm can track the movement of users and obtains optimal downlink capacity. The Prioritized Replay Double DQN algorithm in [14] is applied to make multi-ABS placement decision and improves the coverage rate in complex environment. However, these proposed DRL algorithms are performed in a centralized way, i.e., assume that there is a central controller which can obtain all information and decide which AeBS take what action. As the number of AeBSs increases, the action space grows explosively, bringing huge cost of memory and time, which is impractical for multi-AeBS scenario. But if we directly apply single-agent DRL algorithms to the multi-agent settings by letting each agent independently learn its own Q-function, the environment will become non-stationary from the perspective of any agent. MARL provides a new perspective to address these challenges. In this paper, we modify DQN to extend it to multi-agent environment.

III. System Model

We consider a multi-AeBS communication scenario where N AeBSs support K ground user equipments (UEs), as shown in Fig. 1. The set of AeBSs and UEs are denoted as $\mathbf{N} = \{1, 2, ..., N\}$ and $\mathbf{K} = \{1, 2, ..., K\}$, respectively. The locations of AeBS n and ground user k are expressed as (x_n, y_n, h_n) and $(x_k, y_k, 0)$, respectively.

A. Air-to-Ground Channel Model

The transmission link between AeBSs and UEs is an air-to-ground (A2G) channel. The radio signal transmitted by AeBS first propagates in free space, suffering free space pathloss. When it reaches the urban environment, it will be shadowed and scattered by man-made structure like buildings and trees, which also brings the excessive path loss. Therefore, the A2G mean path loss between AeBS n and ground user k can be modeled as:
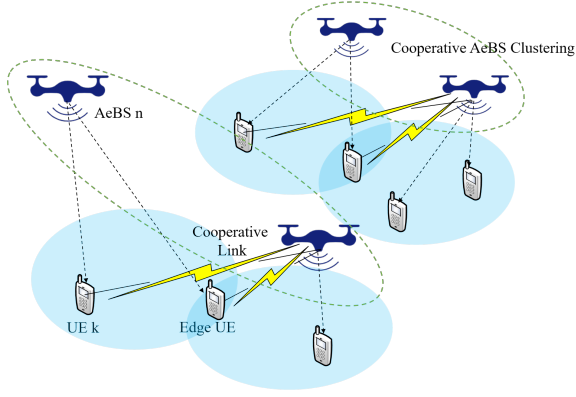
Fig. 1: System model.

$$PL_{LoS}^{n,k} = P_{FS}^{n,k} + \eta_{LoS}, \qquad (1)$$

$$PL_{NLoS}^{n,k} = P_{FS}^{n,k} + \eta_{NLoS}, \qquad (2)$$

where $P_{FS}^{n,k}$ represents the free space pathloss which can be expressed as $P_{FS}^{n,k} = 20log(4\pi f_c d_{n,k}/c)$, where f_c is the carrier frequency, c is the speed of light, and $d_{n,k}$ is the distance between AeBS n and UE k, which can be obtained as $d_{n,k} = \sqrt{(x_n - x_k)^2 + (y_n - y_k)^2 + h_n{}^2}$. η_{LoS} and η_{NLoS} refer to the mean value of the excessive pathloss under the LoS enviornment and non-line-of-sight (NLoS), respectively.

The probability of LoS is considered as a continuous function determined by the environment and elevation angle $\theta_{n,k}$ between AeBS n and UE k.

$$P_{LoS}^{n,k} = \frac{1}{1 + a\exp\left[-b\left(\theta_{n,k} - a\right)\right]}, \qquad (3)$$

where a and b are environment constants. The probability of NLoS can be obtained as $P_{NLoS}^{n,k} = 1 - P_{LoS}^{n,k}$.

Hence, the average pathloss between AeBS n and UE k is:

$$\overline{PL}_{n,k} = P_{LoS}^{n,k} \times PL_{LoS}^{n,k} + P_{NLoS}^{n,k} \times PL_{NLoS}^{n,k} \qquad (4)$$

B. Capacity evaluation model under cooperative scheme

We assume that multiple AeBSs are able to provide services for users collaboratively, forming m cooperative AeBS-UE sets. In m-th cooperative set, the set of AeBSs and UEs are denoted as $\mathbf{N}_m$ and $\mathbf{K}_m$, respectively.

We divide the AeBSs with strong interference into a cooperative set and perform base station cooperation within the set to convert the strong interference into useful signals, and the interference outside the cluster becomes negligible. Therefore, the SINR of the signal received at user k_m, $(k_m \in \mathbf{K}_m)$ can be approximated as:

$$SINR_{k_m} = \frac{\sum P_n \overline{PL}_{n,k}^{-1}}{\sigma^2} \qquad (5)$$

where σ^2 is the noise power.

Then the total throughput of the system can be expressed

as:

$$\varrho = \sum_m \left[\sum_{k_m \in \mathbf{K}_m} B\log_2\left(1 + SINR_{k_m}\right) \right] \qquad (6)$$

where B is channel bandwidth.

C. Problem Formulation

In this paper, our optimization objective is to maximize the throughput of the whole system by adjusting the 3D locations of the cooperative AeBSs and the association of UEs with AeBSs. Therefore, the optimization problem is formulated as follows:

$$\max_{x_n,y_n,h_n,m,\mathbf{N}_m,\mathbf{K}_m} \varrho$$

$$\begin{aligned}
\text{s.t.} \quad &C1: x_{\min} < x_n < x_{\max}, \forall n \in \mathbf{N}\\
&C2: y_{\min} < y_n < y_{\max}, \forall n \in \mathbf{N}\\
&C3: h_{\min} < h_n < h_{\max}, \forall n \in \mathbf{N}\\
&C4: (x_n, y_n, h_n) \neq (x_l, y_l, h_l), \forall n, l \in \mathbf{N}, n \neq l,\\
&C5: \mathbf{N}_1 \cup ... \cup \mathbf{N}_m = \mathbf{N}\\
&C6: \mathbf{K}_1 \cup ... \cup \mathbf{K}_m = \mathbf{K}
\end{aligned} \qquad (7)$$

where ϱ is the system throughput in (6). C1-C3 constraint AeBSs from flying out of the considered region, and C4 represents the collision constraint that prevents AeBSs from flying to the same position. C5 and C6 represent that all users in the system can be served.

IV. 3D DEPLOYMENT OF COOPERATIVE AeBSS FOR CAPACITY MAXIMIZATION

In this section, we propose a two-step mechanism in order to solve the problem presented in (7). Firstly, we divide AeBSs which interfere severely with each other into a cluster according to the interference characteristics. Secondly, we propose a MADQN-based deployment algorithm to decide the positions of AeBSs to achieve maximum system throughput.

A. APC-based cooperation mechanism for AeBSs

The A2G channel between AeBSs and ground users is a LoS-dominated link with inevitably strong cross-link interference, especially in multi-AeBS networks. In the deployment design of the multi-AeBS, many studies adopt a user-clustering approach based on geographical information to simplify the system model, which fails to reflect well the interference characteristics in the actual network. In this paper, we adopt Affinity Propagation (AP) clustering algorithm proposed in [19] for AeBS-UE partitioning, considering the interference characteristics. Specifically, we characterize interference between AeBSs as similarity matrix, and AP algorithm clusters data points based on this matrix. After clustering, AeBSs within one cluster serve their UEs collaboratively, transforming strong interferences to cooperative signals.

Considered a network with N AeBSs, the similarity matrix is denoted as:

$$S = [s_{n,l}]_{N \times N}, \qquad (8)$$

where $s_{n,l}$ indicates the interference of AeBS l to AeBS n associated users K_n, $s_{n,l} = \sum_{k_n \in K_n} P_l \overline{PL}_{l,k_n}$.

AP algorithm treats all data points as potential cluster centers, called *exemplar*. Two types of information are passed between data points in the AP algorithm, *responsability R* and *availability A*. AP clustering is a process of iteratively updating the matrix $R = [r_{n,l}]$ and $A = [a_{n,l}]$. $r_{n,l}$ indicates whether AeBS l is suitable to be the exemplar of AeBS n, and $a_{n,l}$ reflects the suitability of AeBS n to choose AeBS l as its exemplar. The update process of $r_{n,l}$ and $a_{n,l}$ is using the rule of Eq. 1 and Eq. 2 of [19], respectively. The detailed process of our proposed AP clustering-based AeBSs cooperation is illustrated in Algorithm 1.

Algorithm 1: Affinity Propagation Clustering for AeBSs Cooperation

---

1 **Input:** Locations of AeBSs and UEs
2 **Output:** AeBS-UE cooperative set **C**
3 Initialize responsibility $R = [0]_{N \times N}$ and availability $A = [0]_{N \times N}$
4 Get initial AeBS-UE association: UE k chooses the AeBS with the strongest receiving power, ie.,
   $N_k = \operatorname{argmax}_{n \in \mathcal{N}} P_n \overline{PL}_{n,k}$
5 Calculate similarity matrix S according to (8)
6 **for** $iteration \leftarrow 1 : max\_iteration$ **do**
7   Calculate responsibility R and availability A and broadcast
8 **for** *each AeBS* $n \in N$ **do**
9   Get cluster centers:
   $\operatorname{exemplar}(l) = \operatorname{argmax}_{n \in \mathcal{N}} \{a_{n,l} + r_{n,l}\}$

B. MADQN-based 3D Deployment of multi-AeBS for Capacity Maximization

We propose a MADQN-based 3D deployment algorithm with the goal of moving AeBSs to achieve better capacity performance. In our proposed scheme, the basic components are defined as follows:

State: The state of AeBS n is $s_n = \{\mathbf{l}_n, \mathbf{u}, \mathbf{s}\}$, where $\mathbf{l}_n = (x_n, y_n, h_n)$ is the current 3D location of the AeBS n, $\mathbf{u}$ represents the locations of all ground users and $\mathbf{s}$ is the locations of other AeBSs.

Action: The action of AeBS n is moving toward 6 directions or remaining stationary, i.e., $a_n = \{$LEFT, RIGHT, FORWARD, BACKWARD, UP, DOWN, HOVER$\}$.

Reward: The reward is the total system throughput in (6) after each AeBS's position changes at time t.

The goal of each agent is to maximize future rewards by selecting actions through interaction with the environment. Generally, immediate rewards corresponding to the actions selected at each point in time are usually not indicative of a good or bad strategy. Reinforcement Learning assumes that future rewards decay over time with a discount factor γ. Therefore, the sum of future reward at the moment t can be expressed as:

Algorithm 2: MADQN for Multi-AeBS Deployment

---

1 **Initialize:**
2 Replay memory $\mathcal{M}$ with capacity N
3 Online network Q for each AeBS n with random weights θ_n
4 Target network Q' for each AeBS n with weights $\theta'_n = \theta_n$
5 The distribution of ground UEs
6 **for** $episode \leftarrow 1 : max\_episode$ **do**
7   Initializes the locations of AeBSs and get initial state $\mathbf{s}_t^n$
8   **for** $step \leftarrow 1 : max\_step$ **do**
9   **for** $AeBS \quad n \leftarrow 1 : N$ **do**
10   Choose an action a_{t_n} according to ϵ-greedy policy: with probability $1 - \epsilon$, AeBS n selects the action $a_{n_t} = \operatorname{argmax}_{a_{n_t}} Q^*(s_{n_t}, a_{n_t}; \theta_n)$; otherwise, selects a random action.
11   Store the action a_{t_n} into the list $\mathbf{a}_t^n$, which also holds the action selections of the other agents in step t;
12   Execute the actions in the list $\mathbf{a}_t^n$
13   Execute AeBSs clustering according to Algorithm 1 and calculate the throughput under the cooperation scheme.
14   Observe reward $\mathbf{r}_t$ and get the new state $\mathbf{s}_{t+1}^n$;
15   Store $(\mathbf{s}_t^n, \mathbf{a}_t^n, \mathbf{r}_t, \mathbf{s}_{t+1}^n)$ in replay memory $\mathcal{M}$
16   Sample a random minibatch $\mathcal{D}$ from $\mathcal{M}$
17   **for** $AeBS \quad n \leftarrow 1 : N$ **do**
18   Perform a gradient descend step on loss function according to (12) with respect to weight θ_n using data set $\mathcal{D}$
19   Every C steps update target nework for each AeBS n: $\theta'_n \leftarrow \theta_n$

$$R_t = \sum_{t'=t}^{T} \gamma^{t'-t} r_{t'}. \quad (9)$$

For agent n, the maximum action value function $Q^*(s_n, a_n)$ represents the maximum expectation of the reward that can be achieved by a subsequent arbitrary strategy π_n after observing state s_n and executing action a_n, which can be defined as follows:

$$Q^*(s_n, a_n) = \max_{\pi_n} \mathbb{E}[R_t \mid s_t = s_n, a_t = a_n, \pi_n] \quad (10)$$

The above formula can be reduced to Bellman's equation:

$$Q^*(s_n, a_n) = \mathbb{E}\left[r_n + \gamma \max_{a'_n} Q^*(s'_n, a'_n) \mid s_n, a_n\right] \quad (11)$$

Loss function is used to gauge the error between the predicted value and the target value, which can be formulated as:

$$L_n(\theta_n) = \mathbb{E}[y_n - Q(s_n, \ a_n \ ; \ \theta_n)]^2 \quad (12)$$

where $y_n = r + \gamma \cdot \max_{a'_n} Q(s_n, a'_n; \theta'_n)$. The **Q** network needs to be continuously optimized to minimize the loss function in each iteration of the training process. To ensure the computational efficiency, stochastic gradient descent is generally used to optimize the loss function and update the weight parameters. Detailed steps of our proposed MADQN-based multi-AeBS 3D deployment algorithm are shown in

Algorithm 2.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, the performance of our proposed MADQN-based 3D deployment scheme of cooperative AeBSs is evaluated. We consider a region of $6km \times 6km$, in which the ground user terminals are randomly distributed. The vertical flying region of AeBSs is $[100m, 300m]$. The main parameters of network environment are listed in the Table I. Each AeBS has its own online and target network, and each neural network is a 3-layer fully connected neural network that contains a hidden layer with 100 neurons and ReLU activations. The simulations are implemented based on the Python 3.6 and PyTorch 1.8 environment.

TABLE I: Main simulation parameters

Simulation Parameters	Values
Carrier frequency	2500 Hz
Channel bandwidth B	20 MHz
Transmit power of AeBS P_m	30dBm
Environmental constants a, b	9.61, 0.16
Excessive pathloss η_{LoS}, η_{NLoS}	1, 20
Speed of light c	3×10^8 m/s
Thermal noise power density	-174dBm/Hz
Learning rate	0.1
ϵ in ϵ-greedy	0.1
Memory size	10000
Batch size	512
Discount factor	0.9

First of all, we validate if the considered system remains stable by implementing our proposed MADQN algorithm for multi-AeBS cooperative communication in the Fig. 2. In all experiments in this work, we set the number of steps in each episode as 200 and the learning starts after the number of samples of transitions is larger than the capacity of replay memory. Fig. 2a shows the reward achieved per episode during the training stage of the MADQN. The reward curve starts to rise after 50 episodes and gradually converges. Fig. 2b shows that the loss function of our proposed algorithm decreases rapidly and becomes smoother as the training proceeds, which indicates the trained MADQN is convergent.

Fig. 3 plots the 3D positions of the AeBSs and UEs, and different colors characterize the association between AeBSs and UEs. We set the number of AeBSs and ground UEs as 4 and 50, respectively. At initial stage, as shown in the Fig. 3a, AeBSs and UEs are randomly distributed, and each UE chooses AeBS with the strongest signal as its service BS. Fig. 3b and Fig. 3c present the AP clustering and K-means result of initial stage, respectively. As we can see, the K-means-based AeBS-UE partitioning method is simply based on the location information of ground users and AeBS selects UEs which are closer rather than UEs with better channel conditions. In addition, clustering result by K-means is uncertain and is influenced by the initial setted clustering center, while AP clustering can obtain definite clustering results once the interference characteristics are determined. Fig. 3d shows the converged locations of AeBSs and AeBS-UE's association by our proposed scheme. It can be observed

(a)
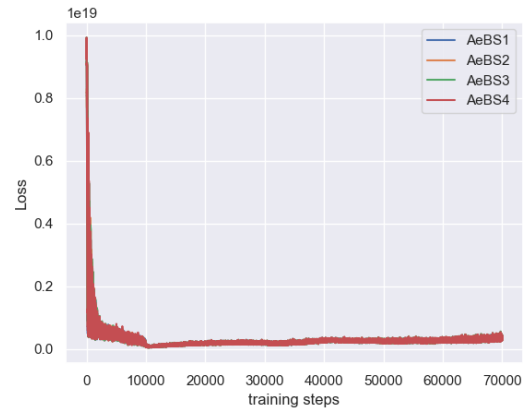
(b)

Fig. 2: Convergence of proposed MADQN. (a): reward versus episodes; (b): loss versus training steps.

that AeBSs are divided into 2 clusters, one of which contains 3 AeBSs providing service collaboratively for the users within the cluster.

To verify the capacity performance improvement brought by the cooperative scheme, we compare it with the "No-CoMP" solution, in which each UE chooses AeBS with the strongest signal as its service BS, without the process of clustering and cooperation. Fig. 4 shows that the proposed cooperative mechanism gets a higher system spectral efficiency when the number of AeBSs increases, while the spectral efficiency of "No-CoMP" system decreases instead of increasing due to the increase of introduced interference. Fig. 4 also shows the effect of the number of UEs on the performance. In the proposed cooperative mechanism, additional users can also collaborate with the AeBSs to establish links and convert the interfering signals into collaborative sources, thus improving the system spectral efficiency. While the No-CoMP system fails to support a large number of users, since there is little significant increase in system spectral efficiency as the number of UEs increases.

We compare our proposed MADQN with centralized DQN in computational time and average converged spectral effi-
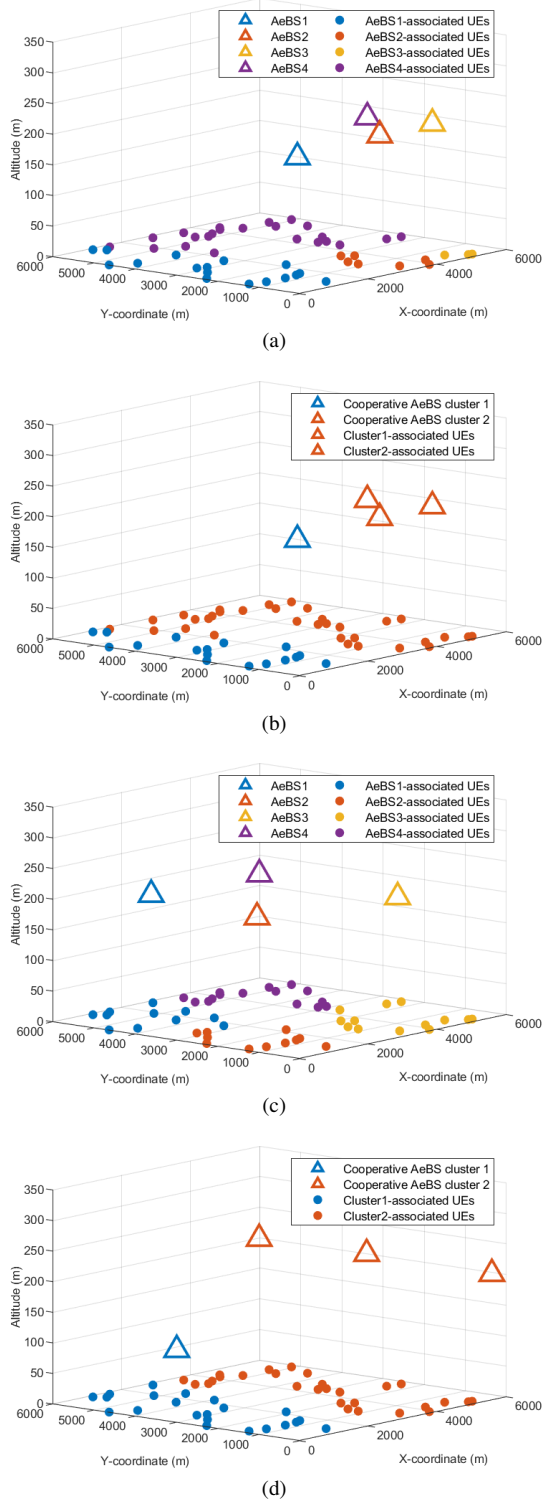
(a)

(b)

(c)

(d)

Fig. 3: 3D distribution and association of AeBSs and ground UEs. (a): Initial stage; (b): Initial stage after AP clustering. (c): Converged stage by K-means algorithm adpoted in [12]. (d): Converged stage by our proposed algorithm.

ciency, as shown in the Fig. 5. We train the MADQN and centralized DQN over 400 training episodes and set the num-
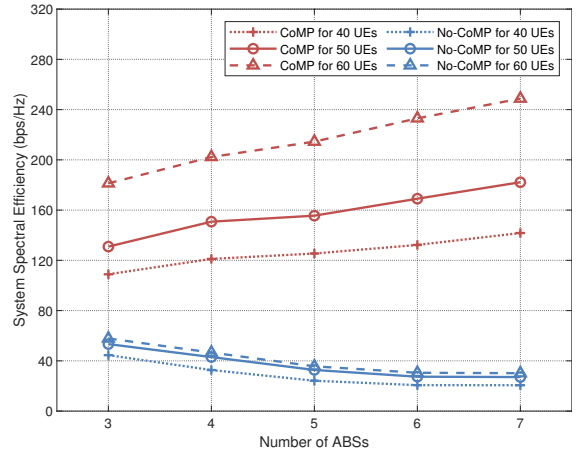

Fig. 4: System spectral efficiency for different numbers of AeBSs.
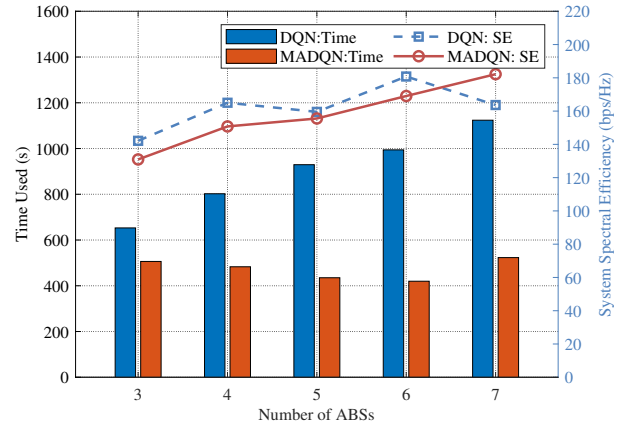

Fig. 5: The performance of different algorithms in computational time and spectral efficiency.

ber of steps in each episode as 200. The two algorithms show little difference in converged . Although the converged system spectral efficiency of DQN is 3.46% higher than MADQN, MADQN tends to be more stable as the number of AeBSs increases. In terms of running time, MADQN consumes less time to train than centralized DQN thanks to its smaller size of action and state space. Additionally, the time required for DQN increases significantly as the number of AeBS increases, while MADQN remains relatively stable.

VI. CONCLUSION

This paper proposes a two-step approach for capacity maximization in the multi-AeBS network, in which AP clustering is used to obtain the cooperation scheme according to the interference characteristics and MADRL is utilized to make deployment decisions of AeBSs. Simulation results demonstrate that the proposed cooperative scheme can greatly improve the network capacity performance since strong interference is eliminated. Result also shows that our proposed MADQN-based multi-AeBS deployment algorithm can achieve higher efficiency in time compared to centralized DQN. In future

work, we will consider the cooperation mechanism between the AeBS-assisted network and the terrestrial network. More factors, such as energy consumption and resource allocation, will also be considered in the formulated optimization problem.

REFERENCES

[1] B. Li, Z. Fei, and Y. Zhang, "UAV communications for 5G and beyond: Recent advances and future trends," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241–2263, 2019.

[2] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[3] R. Shakeri, M. A. Al-Garadi, A. Badawy, A. Mohamed, T. Khattab, A. K. Al-Ali, K. A. Harras, and M. Guizani, "Design challenges of multi-UAV systems in cyber-physical applications: A comprehensive survey and future directions," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3340–3385, 2019.

[4] M. Khoshnevisan, V. Joseph, P. Gupta, F. Meshkati, R. Prakash, and P. Tinnakornsrisuphap, "5G industrial networks with CoMP for URLLC and time sensitive network architecture," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 947–959, 2019.

[5] W. Sun and J. Liu, "Coordinated multipoint-based uplink transmission in internet of things powered by energy harvesting," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2585–2595, 2018.

[6] G. R. MacCartney and T. S. Rappaport, "Millimeter-wave base station diversity for 5G coordinated multipoint (comp) applications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3395–3410, 2019.

[7] J. Lyu, Y. Zeng, R. Zhang, and T. J. Lim, "Placement optimization of UAV-mounted mobile base stations," *IEEE Communications Letters*, vol. 21, no. 3, pp. 604–607, 2017.

[8] C. Qiu, Z. Wei, X. Yuan, Z. Feng, and P. Zhang, "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.

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

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

[11] A. Koushik, F. Hu, and S. Kumar, "Deep Q-learning-based node positioning for throughput-optimal communications in dynamic UAV swarm network," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 554–566, 2019.

[12] P. Yu, G. Jianli, Y. Huo, S. Xiujuan, W. Jiahui, and Y. Ding, "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, 2020.

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

[14] J. Qiu, J. Lyu, and L. Fu, "Placement optimization of aerial base stations with deep reinforcement learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[15] J. Cui, Y. Liu, and A. Nallanathan, "Multi-agent reinforcement learning-based resource allocation for UAV networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 729–743, 2020.

[16] Y. Sun, T. Wang, and S. Wang, "Location optimization and user association for unmanned aerial vehicles assisted mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 056–10 065, 2019.

[17] 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.

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

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