# RESEARCH

## **Open Access**

# Hypergraph convolution mix DDPG for multi-aerial base station deployment



Haoran He<sup>1</sup>, Fanqin Zhou<sup>1</sup>, Yikun Zhao<sup>1</sup>, Wenjing Li<sup>1\*</sup> and Lei Feng<sup>1</sup>

### Abstract

Aerial base stations (AeBS), as crucial components of air-ground integrated networks, can serve as the edge nodes to provide flexible services to ground users. Optimizing the deployment of multiple AeBSs to maximize system energy efficiency is currently a prominent and actively researched topic in the AeBS-assisted edge-cloud computing network. In this paper, we deploy AeBSs using multi-agent deep reinforcement learning (MADRL). We describe the multi-AeBS deployment challenge as a decentralized partially observable Markov decision process (Dec-POMDP), taking into consideration the constrained observation range of AeBSs. The hypergraph convolution mix deep deterministic policy gradient (HCMIX-DDPG) algorithm is designed to maximize the system energy efficiency. The proposed algorithm uses the value decomposition framework to solve the lazy agent problem, and hypergraph convolutional (HGCN) network is introduced to strengthen the cooperative relationship between agents. Simulation results show that the suggested HCMIX-DDPG algorithm outperforms alternative baseline algorithms in the multi-AeBS deployment scenario.

**Keywords** Aerial base station (AeBS), Multi-agent deep reinforcement learning (MADRL), Hypergraph convolution (HGCN), Value decomposition, energy efficiency optimization

#### Introduction

With the exponential growth of data-driven applications and the increasing demand for high-speed and ubiquitous communication services, the integration of edge-cloud computing [1, 2] and aerial wireless networks has become a cornerstone of modern communication systems [3, 4]. The aerial base station (AeBS) [5, 6] is a mobile base station that installs communication equipment on the aerial platform, such as the unmanned aerial vehicle (UAV) [7]. AeBSs can offload computationally intensive tasks to cloud servers, enabling efficient data processing, storage, and analysis. This capability enables AeBSs to provide real-time data services, such as industrial internet of things [8, 9], mobile edge computing networks [10, 11], internet of vehicles [12–14] and health

\*Correspondence:

detection [15, 16]. With the popularity of mobile AeBSs, AeBSs-assisted cloud computing is crucial for delivering diverse services in areas without available infrastructure [17]. To meet the increasing demand for higher communication rates from data-intensive applications and services in future cloud computing, network capacity [18] and system energy consumption [19] have always been the key metrics for network optimization. How to efficiently deploy AeBSs to achieve optimal system energy efficiency is a current research hotspot [20].

In the deployment of AeBSs, traditional heuristic algorithms require repetitive calculations, while deep reinforcement learning (DRL) provides a novel paradigm for enabling the accumulation and utilization of experience in the environment [21, 22]. Recently, the development of computing systems has also led to the widespread application of DRL in wireless communication systems [23]. As the number of AeBSs that need to be controlled grows, researchers are turning to multi-agent deep reinforcement learning (MADRL) for deploying multiple AeBSs. MADRL effectively addresses the challenge of



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Wenjing Li

wjli@bupt.edu.cn

<sup>&</sup>lt;sup>1</sup> State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

dealing with a large action space in centralized singleagent deep reinforcement learning (DRL) when it comes to jointly deploying multiple AeBSs. Additionally, taking into account AeBSs' capability to observe the states of other AeBSs within a certain range, recent studies have employed the decentralized partially observable Markov decision process (Dec-POMDP) [22, 24] to model the deployment problem of AeBSs.

In the cooperative task scenario of partial observable reinforcement learning, for a single agent under the MADRL framework, its reward is likely to be caused by the behavior of its teammates, leading to the training of lazy agents with poor performance[25]. Some researchers proposed value decomposition-based multi-agent reinforcement learning to solve the above lazy agent problem [26]. Through the value decomposition, the contribution of each agent action to the total value function is identified, effectively mitigating the challenges associated with lazy agents in multi-agent reinforcement learning.

For each agent in MADRL, gathering information from other agents is essential for enhancing multi-agent systems coordination. Graph Convolutional Network (GCN) can aggregate neighborhood information and handle the irregular or non-Euclidean nature of graph-structured data [27]. In recent years, some work has introduced GCN into MADRL to enhance the agent cooperation ability and improve the algorithm performance [28, 29]. The simple graph can hardly represent the complex connections between a large number of agents, while hypergraph is a kind of high-dimensional graphic presentation of data, which makes up for the loss of information in the simple graph and is dedicated to describing the system with pairs of combinatorial relations. The properties of the hypergraph allow more information about the nodes contained in the hyperedge. In addition, by selecting a variety of hyperedges, the prior knowledge of the connections between multi-agents can be easily combined, thus enhancing the cooperation between agents [30].

#### **Related work**

Early works adopt heuristic algorithms to solve the deployment problem of AeBSs. Ref. [18] presents an iterative solution that jointly optimizes the AeBS locations and the partially overlapped channel assignment scheme to maximize the throughput of a multi-AeBS system. But the heuristic algorithm is to find the solution of the problem, when the deployment environment of the AeBSs changes slightly, it must be recalculated. The model trained by deep reinforcement learning can make action selections according to the current observation, which effectively solves the problem of repeated calculation of the heuristic algorithm. Ref. [31] proposes a DRL approach that relies on the AeBS networks flow-level

models for learning the optimal traffic-aware AeBS trajectories. Ref. [32] proposes a deep Q-learning algorithm that allows AeBSs to learn the overall network state and account for the joint movement of all AeBSs to adapt their locations.

With the increase of the number of AeBSs to be deployed, the centralized DRL using a single agent will produce the problem of too large action space, and researchers begin to use MADRL. Ref. [33] proposes a MADRL-based approach to minimize the network computation cost while ensuring the quality of service requirements of IoT devices in the UAV-enabled IoT edge network. Ref. [34] proposes an improved clip and countbased proximal policy optimization (PPO) algorithm to solve the partially observable Markov decision process (POMDP) UAV development model. Ref. [35] uses multiagent deep deterministic policy gradient (MADDPG) to maximize the secure capacity by jointly optimizing the trajectory of UAVs and power control. In Refs. [36], MADDPG is adopted to decide the location planning of AeBSs, and results show that the MADDPG-based algorithm is more efficient than centralized DRL algorithms in obtaining the solution.

To solve the problem of cooperation between multiple AeBSs, some researchers have introduced GCN into the deployment of AeBS. Ref. [28] proposes a heterogeneous-graph-based formulation of relations between ground terminals and AeBSs. Ref. [29] proposes a DRLbased control solution to AeBS navigation which enables AeBSs to fly around an unexplored target area under partial observation to provide optimal communication coverage for the ground users. Ref. [37] proposes a GCNbased trajectory planning algorithm that can make AeBSs rebuild communication connectivity during the self-healing process. Ref. [38] proposes a GCN-based MADRL method for UAVs group control, which enables the utilization of mutual interactions among UAVs, resulting in improved signal coverage, fairness, and reduced overall energy consumption.

Motivated by the above literature, this paper introduces the value decomposition algorithm into deep deterministic policy gradient (DDPG) algorithm and adopts hypergraph convolution to learn the cooperative relationship between AeBSs.

#### Motivation

In the face of the high data rate requirements of future services, the energy efficiency enhancement of AeBSbased cloud computing networks is crucial. To achieve this goal, the locations of AeBSs can be adjusted, leverage the advantages of flexible deployment of AeBSs, and adapt to network dynamic requirements. MADRL is regarded as an effective technique for AeBS deployment

due to its better performance compared to centralized DRL. However, MADRL has the following two problems in the deployment of multiple AeBSs: 1) in the MADRL framework, for a single agent, its reward is likely to be caused by the behavior of other agents, leading to the training of lazy agents; 2) it is difficult to learn the cooperative relationship between AeBSs. Aiming at the problem of lazy agents, value decomposition reinforcement learning is used to clarify the contribution of each agent and improve the performance of the multiple AeBSs deployment algorithm. To strengthen the cooperation of AeBSs, the hypergraph convolution is introduced into the deployment of AeBSs, and the cooperation of each AeBS is learned through the hypergraph convolution network. Therefore, in this work, we adopt hypergraph convolution mix DDPG to effectively address the deployment optimization problem of multi-AeBSs.

#### **Our contributions**

In this paper, we study the deployment problem of AeBSs, to maximize the energy efficiency of the multi-AeBS system. We propose the hypergraph convolution mix deep deterministic policy gradient (HCMIX-DDPG) algorithm to solve the optimization problem. The main contributions are listed as follows:

- 1 The energy efficiency maximization problem of AeBSs is modeled as Dec-POMDP in light of the constrained observation range of AeBSs. Furthermore, to combat the issue of lazy agents, we incorporate the concept of value decomposition. This approach helps elucidate the individual contributions of each agent's actions to the overall value function, ultimately enhancing the performance of our multi-AeBS deployment algorithm.
- 2 To strengthen the cooperation among AeBSs, the hypergraph is used to represent the cooperation relationship of each AeBS. The output of the hypergraph convolutional (HGCN) network is the improved value considering the cooperation of each agent. HGCN enhances the cooperation among agents in the MADRL algorithm, and then improves the performance of the algorithm.
- 3 The simulation results show that the performance of HCMIX-DDPG algorithm is better than other baseline DRL algorithms, and the value decomposition and HGCN can improve the deployment performance of multi-AeBS.

#### Organization

The remainder of this paper is structured as follows. In System model section, the system model and problem

formulation are shown. Then, the HCMIX-DDPG algorithm for multi-AeBS deployment is described in Hypergraph convolution mix algorithm for AeBSs section. Discussions of the simulation results are included in Simulation results and discussions section and the paper is finally concluded in Conclusion section.

#### System model

#### System architecture

We consider a multi-AeBS communication scenario where AeBSs serve ground UEs, as shown in Fig. 1. Each AeBS has an observation range, and it can observe the location information of other AeBs within this range. In the current scenario, each AeBs is an agent, which makes decisions based on its state and the state of other AeBss within its observation range.

#### Air-to-ground channel model

AeBSs and user equipments (UEs) communicate data via an air-to-ground (A2G) channel. The mean A2G path loss between UE u and AeBS m can be represented as:

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

$$L_{NLoS}^{m,u} = L_{FS}^{m,u} + \eta_{NLoS},\tag{2}$$

where  $L_{FS}^{m,u}$  represents the free space path loss,  $L_{FS}^{m,u} = 20log(4\pi f_c d_{m,u}/c)$ , where *c* is the light speed,  $f_c$  is the carrier frequency, and  $d_{m,u}$  is the distance between AeBS *m* and UE *u*.  $\eta_{LoS}$  and  $\eta_{NLoS}$  refer to the mean excessive path loss under the LoS and NLoS environment, respectively.

The probability of LoS is related to the communication environment constants  $\alpha$ ,  $\beta$  and elevation angle  $\theta_{m,u}$ , which can be expressed as [39]:

$$P_{LoS}^{m,u} = \frac{1}{1 + \alpha \exp\left[-\beta\left(\theta_{m,u} - \alpha\right)\right]}.$$
(3)

And the probability of NLoS can be obtained as  $P_{NLoS}^{m,u} = 1 - P_{LoS}^{m,u}$ .

The average path loss between UE *u* and AeBS *m* is:

$$L_{m,u} = P_{LoS}^{m,u} \times L_{LoS}^{m,u} + P_{NLoS}^{m,u} \times L_{NLoS}^{m,u}.$$
 (4)

We consider AeBSs to serve ground users through millimeter wave beams. Therefore, the directional mmWave antenna gain also plays a significant role in the AeBS channel in addition to the A2G propagation path loss. In this work, we adopt the 3D mmWave beam scheduling model in Ref. [40].



Fig. 1 System Architecture.

#### **Capacity model**

Each UE is associated with the unique AeBS with the strongest received signal, and the signal interference of other AeBSs to this UE is considered. If AeBS *m* and UE *u* are associated, the SINR  $\xi_u$  of the signal received at UE *u* can be expressed as:

$$\xi_{u} = \frac{P_{m}G_{T}G_{R}L_{m,u}^{-1}}{\sum_{i \in M(i \neq m)} P_{i}G_{T}G_{R}L_{i,u}^{-1} + \sigma^{2}}, u \in U,$$
(5)

where  $P_m$  is the transmit power of AeBS m, M is the total number of AeBSs in the system,  $\sigma^2$  is the thermal noise power,  $G_T$  and  $G_R$  are the main lobe gain of the transmitter and receiver. In this paper, each UE selects the AeBS with the largest SINR for the association.

The capacity of UE *u* can be expressed as:

$$\varrho_u = \eta_u \left( 1 - \frac{\tau}{T} \right) B \log_2 \left( 1 + \xi_u \right), \tag{6}$$

where  $\tau$  is the beam alignment time, *B* represents the channel bandwidth, and *T* is the time slot,  $\eta_u$  is the average ratio of time-frequency resources that can be calculated from Eq.(7) in Ref. [40],  $\eta_u$  occupied by UE *u* is given by:

$$\eta_u = \begin{cases} 1, N_b \le N_u, \\ \frac{N_b}{N_u}, N_b > N_u, \end{cases}$$
(7)

where  $N_b$  is the number of mmWave beams of the AeBS and  $N_u$  is the number of UEs served by this AeBS.

Then the system capacity is as follows:

$$\varrho = \sum_{m=1}^{M} \left[ \sum_{u \in U_m} \varrho_u \right].$$
(8)

where  $U_m$  is users served by AeBS m.

#### **Energy consumption model**

The system's energy consumption is categorized into two components: the first component is the power consumption for propelling the AeBSs, and the second component is the energy consumed for communication by the AeBSs serving UEs.

The propulsion power consumption  $P_w$  [41] of the AeBS can be modeled as:

$$P_{w} = P_{0} \left( 1 + \frac{3V^{2}}{U_{t}^{2}} \right) + P_{1} \left( \sqrt{1 + \frac{V^{4}}{4v_{0}^{4}}} - \frac{V^{2}}{2v_{0}^{2}} \right)^{1/2} + \frac{1}{2} d_{0} \rho s A V^{3},$$
(9)

where  $P_0$  and  $P_1$  represent the blade profile power and induced power during AeBS hovering,  $v_0$  is the average rotor-induced velocity at hover,  $U_t$  is the rotor blade tip speed,  $d_0$  is the fuselage drag ratio, *s* is the rotor solidity, *A* denotes the rotor disc area, and  $\rho$  stands for air density.

The communication power consumption  $P_c$  of the AeBS can be modeled as:

$$P_c = \sum_{u \in \mathcal{U}_m} P_{m,u},\tag{10}$$

where  $P_{m,u}$  is the transmitting power of AeBS m to UE u. The total energy consumption definition of the system is modeled as follows:

$$P_{tot} = \sum_{i=1}^{n} \left[ P_{w,i} + P_{c,i} \right].$$
(11)

where  $P_{w,i}$  and  $P_{c,i}$  the propulsion power consumption and communication power consumption of the *i*-th AeBS, *n* is the number of AeBS.

#### **Problem formulation**

Our goal in this work is to maximize the energy efficiency of the entire system. The energy efficiency of the system  $J_{tot}$  is defined as follows:

$$J_{tot} = \frac{\varrho}{P_{tot}}.$$
 (12)

Consequently, the formulation of the optimization problem is as follows:

$$\max_{(x_m, y_m, h_m)} J_{tot}$$
s.t. C1 :  $x_m \in [x_{\min}, x_{\max}], \forall m \in M,$   
C2 :  $y_m \in [y_{\min}, y_{\max}], \forall m \in M,$   
C3 :  $h_m \in [h_{\min}, h_{\max}], \forall m \in M,$   
C4 :  $(x_m, y_m, h_m) \neq (x_l, y_l, h_l),$   
 $\forall m, l \in M, m \neq l,$   
C5 :  $V_m \leq V_{max}, \forall m \in M,$ 
(13)

where  $J_{tot}$  energy efficiency in (12). C1-C3 constrain AeBSs from traveling out of the considered region, and C4 is the collision restriction. C5 ensures that each AeBS does not move faster than the maximum speed  $V_{max}$ .

#### Hypergraph convolution mix algorithm for AeBSs

In this section, we propose the HCMIX-DDPG algorithm to decide the positions of AeBSs to achieve maximum system energy efficiency. In HCMIX-DDPG, each agent applies a deep deterministic policy gradient (DDPG) [42] to learn individual action value. DDPG is divided into two networks, the actor network and the critric network, the actor network generates the actions of the agent, and the critic network evaluates the action and generation value of the action. The structure of HCMIX-DDPG is presented in Fig. 2. The input of HGCN network is the observation of each agent  $o_m$  and the Q value of its action  $Q(o_m)$ . The improved Q value  $Q'(o_m)$  of HGCN network output and the multi-agent global state s are input into value mixing network and the total value  $Q_{tot}(s)$  of joint action is obtained.



 $Q_{tot}(s)$ 

S

**Reinforcement learning framework** 

Fig. 2 The structure of HCMIX-DDPG

 $(o_1, a_1)$ 

Given that the multi-AeBS deployment involves a continuous action space, we employ the Deep Deterministic Policy Gradient (DDPG) framework for each agent. The AeBS is the agent in our HCMIX-DDPG algorithm, and its observation, action, and reward are specified as follows:

 $(o_m, a_m)$ 

**Observation:** Each AeBS has observaan tion range, the observation of each AeBS is the current position of the AeBS and the posiobservation tion of other AeBSs within its range  $o_m = (o_{1xm}, o_{1ym}, o_{2xm}, o_{2ym}, ..., o_{nxm}, o_{nym}),$  $o_m \in O, m \in \{1, 2, ..., n\}$ , where  $o_{1xm}$ ,  $o_{1ym}$  respectively represent the X-axis position and Y-axis position of the AeBS itself. Other observed values represent the X-axis position and Y-axis position of other AeBSs. When other AeBSs are not in the observation range, the corresponding observation value is 0.

Action: Each agent's action is defined as the distance traveled in the X-axis and Y-axis  $a_m = \{(\Delta x_m, \Delta y_m)\}, a_m \in A, m \in \{1, 2, ..., n\}.$ 

**Reward:** As a fully cooperative reinforcement learning scenario, the reward is the system energy efficiency in E.q. (14).

#### Hypergraph convolutional network

Hypergraph is a generalized graph structure, which can describe the relationship and constraints between nodes more flexibly.

According to Ref. [43], HGCN formula is defined as follows:

$$\boldsymbol{x}^{(l+1)} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{H} \boldsymbol{W} \boldsymbol{B}^{-1} \boldsymbol{H}^{T} \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{x}^{(l)} \boldsymbol{P}, \qquad (14)$$

where  $\mathbf{x}^{(l+1)}$  is the output of the hypergraph convolution,  $\mathbf{x}^{(l)}$  is the input of the hypergraph convolution, H is the adjacency matrix of the hypergraph, W is the weight matrix of hyperedges, B is the hyperedge degree matrix, D is the vertex degree matrix of the hypergrap and Pdenotes the weight matrix between the (l)-th and (l + 1)-th layer.

According to Ref. [30], we build the adjacency matrix H of the hypergraph. As shown in the following equation, O is taken as input to generate the first part of the adjacency matrix of the hypergraph  $H_1$ :

$$H_1 = \text{ReLU}(\text{Linear}(O)). \tag{15}$$

The final hypergraph adjacency matrix H is defined as follows:

$$H = \begin{bmatrix} \omega_{1,1}, \cdots, \omega_{1,m} \\ \omega_{2,1}, \cdots, \omega_{2,m} \\ \vdots, & \ddots, & \vdots \\ \omega_{n,1}, & \cdots, & \omega_{n,m} \end{bmatrix} \begin{bmatrix} H_a, \cdots, & 0 \\ 0, & \cdots, & 0 \\ \vdots, & \ddots, & \vdots \\ 0, & \cdots, & H_a \end{bmatrix}, \quad (16)$$

where  $w_{i,k}$  represents the weight of learned connection between agents, *n* refers to the number of agents, *m* refers to the number of hyperedges, and  $H_a$  signifies the average value of  $H_1$ .

#### **Mixing network**

The adjacency matrix of the hypergraph is constructed by the observations of all agents according to E.q. (15), (16), and the independent action value of each agent is substituted into E.q. (14) to obtain the improvement value of the cooperation relationship of all agents:

$$Q' = \text{HGCN}(\text{HGCN}(Q, \text{abs}(V_1)), \text{abs}(V_2)), \quad (17)$$

where  $V_i$  is the weight matrix in the *i*-th layer and Q denotes the improvement value.

We employ value function decomposition in the multiagent actor critique framework [44]. The QMIX-DDPG module uses a deep neural network to aggregate the values of each agent into a total value.  $Q_{tot}$  stands for total value of joint actions, which can be formulated as:

$$Q_{tot} = f_s(Q_1(s, a_1), \dots, Q_n(s, a_n)), \frac{\partial f_s}{\partial Q_a} \ge 0,$$
(18)

where *s* is the global state,  $a_n$  is the action of the *n*-th agent.

The proposed HCMIX-DDPG algorithm for AeBS deployment is shown in Algorithm 1.

1: Initialize:
2: The actor network and the critic network for each AeBS
3: Hypergraph convolutional network Mixing network
4: for episode = 1 : max_episode do
5: Initialize the locations of AeBSs
6: Initialize the locations of UEs
7: for $step = 1 : max\_step$ do
8: Each AeBS m obtains its observation om
9: The actor network generates action $a_m$ and the critic network evaluates the value $Q(o_m)$ of
this action
10: Through the HGCN network, the value $Q^{i}(o_{m})$ considering the cooperation relationship of
the agents is generated
11: Through the mixing network, $Q_{i}(o_{m})$ is aggregated into $Q_{tot}$
12: Each AeBS observes reward $r_m$ , executes action $a_m$ and updates observation $o_m$
13: Store transition $(o_m, a_m, r_m, o_{m+1})$ in D
14: Sample random minibatch of transitions and training
15: Soft update the actor network and the critic network
16: end for
17: end for

Algorithm 1 HCMIX-DDPG algorithm for AeBS deploymentOur proposed HCMIX-DDPG adopts the centralized training decentralized execution (CTDE) deployment mode [45]. The HCMIX-DDPG is divided into two phases, the training phase and the execution phase. In the training phase, strong computing power resources are needed to train all models, which can use the computing power of cloud technology. After fully trained, each actor network is deployed to the corresponding AeBS, and the AeBS makes a decision based on its observations to complete the deployment. The CTDE framework is especially suitable for AeBS-assisted networks since it brings less communication burden to resource-limited AeBSs [46]. The overhead of the execution phase is analyzed in the following, we assume that an actor network has  $N_h^a$  hidden layers, which have  $q_i^a$  neurons respectively,  $i = 1, ..., N_h^a$ . The time complexity of HCMIX-DDPG in each iteration can be obtained as  $O\left(\sum_{i=0}^{N_h^a-1} q_i^a q_{i+1}^a\right)$  [47]. After the actor networks are deployed on the AeBSs, the AeBSs can be commanded to move according to the real-time observation, and finally the deployment task can be completed.

#### Simulation results and discussions

Simulation is carried out in a 2 km × 2 km environment. The maximum length of each step in the direction  $(x_m, y_m)$  is 100 m, the flight altitude of AeBSs is 20 m, and the observation range of AeBSs is 200 m. The simulation parameters of the environment [40, 48] are listed in Table 1 and the hyperparameters of the DDPG are listed in Table 2. According to [49], the environmental parameters for the three different channel conditions are listed in Table 3.

We compare our proposed HCMIX-DDPG with the other three DRL algorithms: Value-Decomposition

Simulation Parameters	Values
Speed of light <i>c</i>	3 × 10 <sup>8</sup> m/s
Channel bandwidth B	150 MHz
Carrier frequency $f_c$	30 GHz
Pilot duration ratio $T_p/T$	$2 \times 10^{-4}$
Thermal noise power density	-174 dBm/Hz
Tip speed of the blade $U_t$	60 m/s
Mean induced speed $v_0$	2.486 m/s
lade profile power $P_0$	112.203 W
induced power $P_1$	10.574 W
Fuselage drag proportion $d_0$	0.5
Air density $ ho$	1.225 kg/m <sup>3</sup>
Rotor solidity s	0.0832
Rotor disc area A	0.2827 m <sup>2</sup>

 Table 2
 The hyperparameters of the HCMIX-DDPG

Network hyperparameters	Values
Layers of actor, critic	4, 4
Neurons of hidden layers for actor	[64,64]
Neurons of hidden layers for critic	[64,64]
Learning rate of actor, critic	0.01, 0.01
Optimizer	Adam
Activation Function	ReLU
Soft updating rate	0.5
$\epsilon$ in $\epsilon$ -greedy	0.1
Memory size	10000
Batch size	256

 Table 3
 Channel environment parameters

Channel environment	$(\alpha, \beta, \eta_{LOS}, \eta_{NLOS})$
Suburban	(4.88 0.43,0.1,21)
Urban	(9.61 0.16,1,20)
High-rise urban	(27.23,0.08,2.3,34)

Network DDPG (VDN-DDPG), QMIX-DDPG and independent DDPG (IDDPG).

The agents of four deep reinforcement learning algorithms all adopt the DDPG framework, and the details of the four algorithms are as follows:

 HCMIX-DDPG: In HCMIX-DDPG, the HGCN is introduced into the value decomposition reinforcement learning framework to strengthen the learning of the cooperation relationship of each agent, and the QMIX-DDPG network aggregates the total value.

- 2. QMIX-DDPG: QMIX [44] is a classical value decomposition reinforcement learning framework, which uses a deep neural network to combine the individual values of each agent into a total value.
- 3. VDN-DDPG: VDN [26] is a classical value decomposition reinforcement learning framework, which uses a linear function to combine the values of each agent into a total value.
- 4. IDDPG [50]: IDDPG is a fully distributed algorithm. In IDDPG, each AeBS makes decisions based on its observation.

Figure 3 illustrates how network performance is affected by varying the number of UEs. By comparing our proposed scheme with the other three DRL algorithms, it can be found that the IDDPG has the worst performance in energy efficiency. This is because the multi-AeBS deployment algorithm performs poorly as a result of the lazy agent problem, which is caused by the fact that IDDPG does not introduce the concept of value decomposition. While QMIX-DDPG uses a deep neural network to aggregate the values of each agent into a total value, VDN-DDPG makes use of a linear function to do so. HCMIX-DDPG adds a hypergraph convolution module based on QMIX-DDPG, and uses the HGCN module to generate the improved value considering the cooperation relationship of agents when generating the value of each agent. Simulation results show that the algorithm of HCMIX-DDPG has the best performance in energy efficiency, and the algorithm of QMIX-DDPG has better performance than VDN-DDPG in the case of different UEs. And by comparing HCMIX-DDPG, QMIX-DDPG, it can be found that the energy consumption of these three algorithms is similar, but HCMIX-DDPG improves the capacity and then improves the energy efficiency of the system.

Figure 4 illustrates how network performance is affected by varying the number of AeBSs. It can be found that with the increase of AeBSs, the energy efficiency, capacity, and energy consumption of the system show an increasing trend. Although the increase in the number of AeBS will indicate the total energy consumption of the system, it can better serve the users within the system and improve the total system capacity, thus improving the energy efficiency of the system. By comparing Figs. 3 and 4, we can find that the algorithm performance of HCMIX-DDPG is better than that of other algorithms. In the case of different numbers of AeBSs and different numbers of UEs, HCMIX-DDPG is better than other algorithms in energy efficiency and capacity.



Fig. 3 Comparison of network performance metrics in different numbers of UEs. **a** Energy efficiency **b** Capacity **c** Energy consumption

HCMIX-DDPG a QMIX-DDPG 1.8 VDN-DDPG ▼-- IDDPG Energy efficiency (Mbit/Joule) 1.7 1.6 1.5 1.3 1.2 10 12 14 8 Number of AeBSs ×10 1.7 HCMIX-DDPG QMIX-DDPG 1.6 VDN-DDPG 1.5 1. (Mbit/s) 1.3 Capacity 1.2 0.9 0.8 12 14 8 10 Number of AeBSs 9500 -- HCMIX-DDPG OMIX-DDPG VDN-DDPG 9000 - IDDPG consumption (Joule) 8500 8000 C float 7500 Euergy 7000 6500 8 10 12 14 Number of AeBSs

1.9

Fig. 4 Comparison of network performance metrics in different numbers of AeBSs. a Energy efficiency b Capacity c Energy consumption

Figure 5 shows the energy efficiency of the system for different numbers of AeBSs under different channel conditions. It can be found that the system energy efficiency is the highest in suburban, followed by urban and the lowest in high-rise urban. This is due to the fact that for AeBSs, the suburban areas have better channel



**Fig. 5** Energy efficiency of the system for different numbers of AeBS under different channel conditions

conditions while there are many blockages in the highrise urban environment, which affects the system capacity and thus the energy efficiency. And with the increase in the number of AeBSs, the energy efficiency of the system gradually increases.

Figure 6 shows energy efficiency with different numbers of mmWave beams of the AeBS. Numbers of mmWave beams is  $N_b$  in Eq. (7). It can be found that when the number of beams increases, the system energy efficiency is higher, which is due to the increase of communication resources of the system.

Figure 7 shows energy efficiency under different mean values of transmit power. In the simulation, the transmit power of each AeBS obeys a normal distribution with a standard deviation of 2. It can be found from the figure that the system energy efficiency decreases as the mean value of AeBS transmit power increases. Although



Fig. 6 Energy efficiency with different numbers of mmWave beams



Fig. 7 Energy efficiency under different mean value of transmit power

increasing the transmission power can increase the system capacity, it will also increase the communication energy consumption in the energy consumption, which will lead to a decrease of energy efficiency.

#### Conclusion

This paper models the deployment problem of multi-AeBS as a Dec-POMDP and adopts the MADRL framework to maximize the energy efficiency of the system. An algorithm called HCMIX-DDPG is designed to solve the formulated problem, which combines the value decomposition framework with HGCN, aiming to address the lazy agent problem and strengthen the cooperative relationship between agents in the MADRL settings. The performance of our suggested HCMIX-DDPG algorithm, which can increase the energy efficiency of multi-AeBS systems and improve the efficiency of joint deployment of multi-AeBS, is superior to existing algorithms, according to simulation findings.

#### Authors' contributions

Haoran He was responsible for data visualization and prepared the analytical figures and the technical draft of the manuscript. Fanqin Zhou guided the design of the algorithms and experiment and prepared the final manuscript for submission. Yikun Zhao designed the algorithm, helped with setting up the experiment environment, and refined the whole text of the manuscript. Wenjing Li secured the research funding and conducted an investigation into the manuscript's background. Lei Feng refined the whole text of the manuscript and helped with preparing the final manuscript for submission. All the authors reviewed the manuscript.

#### Funding

This research was supported by the National Natural Science Foundation of China (No. 61971053).

#### Availability of data and materials

Not applicable.

#### Declarations

Ethics approval and consent to participate Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 11 September 2023 Accepted: 25 November 2023 Published online: 06 December 2023

#### References

- Wang F, Li G, Wang Y, Rafique W, Khosravi MR, Liu G, Liu Y, Qi L (2023) Privacy-aware traffic flow prediction based on multi-party sensor data with zero trust in smart city. ACM Trans Internet Technol 23(3):1–19
- Jia Y, Liu B, Dou W, Xu X, Zhou X, Qi L, Yan Z (2022) Croapp: a CNN-based resource optimization approach in edge computing environment. IEEE Trans Ind Inf 18(9):6300–6307
- Pham QV, Ruby R, Fang F, Nguyen DC, Yang Z, Le M, Ding Z, Hwang WJ (2022) Aerial computing: A new computing paradigm, applications, and challenges. IEEE Internet Things J 9(11):8339–8363. https://doi.org/10. 1109/JIOT.2022.3160691
- Gong Y, Chen K, Niu T, Liu Y (2022) Grid-based coverage path planning with NFZ avoidance for UAV using parallel self-adaptive ant colony optimization algorithm in cloud IoT. J Cloud Comput 11(1):29
- 5. Lai CC, Tsai AH, Ting CW, Lin KH, Ling JC, Tsai CE (2023) Interferenceaware deployment for maximizing user satisfaction in multi-UAV wireless networks. IEEE Wirel Commun Lett 12(7):1189–1193
- Mao H, Liu Y, Xiao Z, Han Z, Xia XG (2023) Joint resource allocation and 3D deployment for multi-UAV covert communications. IEEE Internet Things J
- Qu C, Sorbelli FB, Singh R, Calyam P, Das SK (2023) Environmentally-aware and energy-efficient multi-drone coordination and networking for disaster response. IEEE Trans Netw Serv Manag 20(2):1093–1109
- Yang Y, Yang X, Heidari M, Khan MA, Srivastava G, Khosravi M et al (2022) ASTREAM: data-stream-driven scalable anomaly detection with accuracy guarantee in IIoT environment. IEEE Trans Netw Sci Eng 10(5):3007–3016
- Qi L, Yang Y, Zhou X, Rafique W, Ma J (2021) Fast anomaly identification based on multiaspect data streams for intelligent intrusion detection toward secure Industry 4.0. IEEE Trans Ind Inf 18(9):6503–6511
- Wang W, Srivastava G, Lin JCW, Yang Y, Alazab M, Gadekallu TR (2022) Data freshness optimization under CAA in the UAV-aided MECN: a potential game perspective. IEEE Trans Intell Transp Syst 24(11):12912–12921
- 11. Yang Y, Wei X, Xu R, Peng L (2021) Joint optimization of Aol, SINR, completeness, and energy in UAV-aided SDCNs: Coalition formation game and cooperative order. IEEE Trans Green Commun Netw 6(1):265–280
- Xu X, Jiang Q, Zhang P, Cao X, Khosravi MR, Alex LT, Qi L, Dou W (2022) Game theory for distributed IoV task offloading with fuzzy neural network in edge computing. IEEE Trans Fuzzy Syst 30(11):4593–4604
- Xu X, Fang Z, Zhang J, He Q, Yu D, Qi L, Dou W (2021) Edge content caching with deep spatiotemporal residual network for IoV in smart city. ACM Trans Sensor Netw (TOSN) 17(3):1–33
- Yang Y, Wang W, Liu L, Dev K, Qureshi NMF (2022) Aol optimization in the UAV-aided traffic monitoring network under attack: A stackelberg game viewpoint. IEEE Trans Intell Transp Syst 24(1):932–941
- Kong L, Wang L, Gong W, Yan C, Duan Y, Qi L (2022) LSH-aware multitype health data prediction with privacy preservation in edge environment. World Wide Web 25(5):1793–1808
- Xu X, Tian H, Zhang X, Qi L, He Q, Dou W (2022) Discov: Distributed COVID-19 detection on X-ray images with edge-cloud collaboration. IEEE Trans Serv Comput 15(3):1206–1219
- Zhou Y, Ge H, Ma B, Zhang S, Huang J (2022) Collaborative task offloading and resource allocation with hybrid energy supply for UAV-assisted multiclouds. J Cloud Comput 11(1):42
- Zou C, Li X, Liu X, Zhang M (2021) 3D placement of unmanned aerial vehicles and partially overlapped channel assignment for throughput maximization. Digit Commun Netw 7(2):214–222

- Deng D, Li X, Menon V, Piran MJ, Chen H, Jan MA (2022) Learning-based joint UAV trajectory and power allocation optimization for secure IoT networks. Digit Commun Netw 8(4):415–421
- Dai B, Niu J, Ren T, Hu Z, Atiquzzaman M (2021) Towards energy-efficient scheduling of UAV and base station hybrid enabled mobile edge computing. IEEE Trans Veh Technol 71(1):915–930
- Liu R, Qu Z, Huang G, Dong M, Wang T, Zhang S et al (2022) DRL-UTPS: DRL-based trajectory planning for unmanned aerial vehicles for data collection in dynamic IoT network. IEEE Trans Intell Veh 8(2):1204–1218
- 22. Wang Y, Wang J, Zhang W, Zhan Y, Guo S, Zheng Q, Wang X (2022) A survey on deploying mobile deep learning applications: A systemic and technical perspective. Digit Commun Netw 8(1):1–17
- Dai H, Yu J, Li M, Wang W, Liu AX, Ma J, Qi L, Chen G (2023) Bloom filter with noisy coding framework for multi-set membership testing. IEEE Trans Knowl Data Eng 35(7):6710–6724. https://doi.org/10.1109/TKDE. 2022.3199646
- 24. Yin S, Yu FR (2021) Resource allocation and trajectory design in UAV-aided cellular networks based on multiagent reinforcement learning. IEEE Internet Things J 9(4):2933–2943
- Sharma PK, Fernandez R, Zaroukian E, Dorothy M, Basak A, Asher DE (2021) Survey of recent multi-agent reinforcement learning algorithms utilizing centralized training. In: Artificial intelligence and machine learning for multi-domain operations applications III, vol 11746. SPIE, Bellingham, p. 665–676
- Li J, Kuang K, Wang B, Liu F, Chen L, Fan C et al (2022) Deconfounded value decomposition for multi-agent reinforcement learning. In: Proceedings of the 39th International Conference on Machine Learning. JMLR, Cambridge, p. 12843–12856
- Hossain RR, Huang Q, Huang R (2021) Graph convolutional networkbased topology embedded deep reinforcement learning for voltage stability control. IEEE Trans Power Syst 36(5):4848–4851
- Zhang X, Zhao H, Wei J, Yan C, Xiong J, Liu X (2022) Cooperative trajectory design of multiple UAV base stations with heterogeneous graph neural networks. IEEE Trans Wirel Commun 22(3):1495–1509
- Ye Z, Wang K, Chen Y, Jiang X, Song G (2022) Multi-UAV navigation for partially observable communication coverage by graph reinforcement learning. IEEE Trans Mob Comput 22(7):4056–4069
- Bai Y, Gong C, Zhang B, Fan G, Hou X, Lu Y (2022) Cooperative multi-agent reinforcement learning with hypergraph convolution. In: 2022 International Joint Conference on Neural Networks (IJCNN). IEEE, Piscataway, NJ, p. 1–8
- Saxena V, Jaldén J, Klessig H (2019) Optimal UAV base station trajectories using flow-level models for reinforcement learning. IEEE Trans Cogn Commun Netw 5(4):1101–1112
- Luong P, Gagnon F, Tran LN, Labeau F (2021) Deep reinforcement learning-based resource allocation in cooperative UAV-assisted wireless networks. IEEE Trans Wirel Commun 20(11):7610–7625. https://doi.org/10. 1109/TWC.2021.3086503
- Seid AM, Boateng GO, Mareri B, Sun G, Jiang W (2021) Multi-agent DRL for task offloading and resource allocation in multi-UAV enabled IoT edge network. IEEE Trans Netw Serv Manag 18(4):4531–4547
- 34. Dai C, Zhu K, Hossain E (2022) Multi-agent deep reinforcement learning for joint decoupled user association and trajectory design in full-duplex multi-UAV networks. IEEE Trans Mob Comput 22(10):6056–6070
- Zhang Y, Mou Z, Gao F, Jiang J, Ding R, Han Z (2020) UAV-enabled secure communications by multi-agent deep reinforcement learning. IEEE Trans Veh Technol 69(10):11599–11611
- Wang W, Lin Y (2021) Trajectory design and bandwidth assignment for UAVs-enabled communication network with multi-agent deep reinforcement learning. In: 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall). IEEE, Piscataway, NJ, p. 1–6
- Mou Z, Gao F, Liu J, Wu Q (2021) Resilient UAV swarm communications with graph convolutional neural network. IEEE J Sel Areas Commun 40(1):393–411
- Dai A, Li R, Zhao Z, Zhang H (2020) Graph convolutional multi-agent reinforcement learning for UAV coverage control. In: 2020 International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, Piscataway, NJ, p. 1106–1111
- Al-Hourani A, Kandeepan S, Lardner S (2014) Optimal LAP altitude for maximum coverage. IEEE Wirel Commun Lett 3(6):569–572. https://doi. org/10.1109/LWC.2014.2342736

- Zhao Y, Zhou F, Feng L, Li W, Yu P (2023) MADRL-based 3D deployment and user association of cooperative mmWave aerial base stations for capacity enhancement. Chin J Electron 32(2):283–294
- 41. Zeng Y, Xu J, Zhang R (2019) Energy minimization for wireless communication with rotary-wing UAV. IEEE Trans Wirel Commun 18(4):2329–2345
- 42. Yu Y, Tang J, Huang J, Zhang X, So DKC, Wong KK (2021) Multi-objective optimization for UAV-assisted wireless powered IoT networks based on extended DDPG algorithm. IEEE Trans Commun 69(9):6361–6374
- Bai S, Zhang F, Torr PH (2021) Hypergraph convolution and hypergraph attention. Pattern Recog 110:107637
- Su J, Adams S, Beling P (2021) Value-decomposition multi-agent actorcritics. In: Proceedings of the AAAI conference on artificial intelligence, vol 35. AAAI, Menlo Park, CA, p. 11352–11360
- Azzam R, Boiko I, Zweiri Y (2023) Swarm cooperative navigation using centralized training and decentralized execution. Drones 7(3):193
- Hwang S, Lee H, Park J, Lee I (2022) Decentralized computation offloading with cooperative UAVs: Multi-agent deep reinforcement learning perspective. IEEE Wirel Commun 29(4):24–31
- Gao A, Wang Q, Liang W, Ding Z (2021) Game combined multi-agent reinforcement learning approach for UAV assisted offloading. IEEE Trans Veh Technol 70(12):12888–12901
- Zeng Y, Zhang R (2017) Energy-efficient UAV communication with trajectory optimization. IEEE Trans Wirel Commun 16(6):3747–3760
- Diao D, Wang B, Cao K, Dong R, Cheng T (2022) Enhancing reliability and security of UAV-enabled NOMA communications with power allocation and aerial jamming. IEEE Trans Veh Technol 71(8):8662–8674
- Ciftler BS, Alwarafy A, Abdallah M (2021) Distributed DRL-based downlink power allocation for hybrid RF/VLC networks. IEEE Photonics J 14(3):1–10

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com