Multi-Agent Collaborative Learning for UAV Enabled Wireless Networks

Wenchao Xia, *Member, IEEE*, Yongxu Zhu, *Senior Member, IEEE*, Lorenzo De Simone, Tasos Dagiuklas, *Member, IEEE*, Kai-Kit Wong, *Fellow, IEEE*, Gan Zheng, *Fellow, IEEE*

***Abstract*—The unmanned aerial vehicle (UAV) technique provides a potential solution to scalable wireless edge networks. This paper uses two UAVs, with accelerated motions and fixed altitudes, to realize a wireless edge network, where one UAV forwards downlink signals to user terminals (UTs) distributed over an area while the other one collects uplink data. The conditional average achievable rates, as well as their lower bounds, of both the uplink and downlink transmission are derived considering the active probability of UTs and the service queues of two UAVs. In addition, a problem aiming to maximize the energy efficiency of the whole system is formulated, which takes into account communication related energy and propulsion energy consumption. Then, we develop a novel multi-agent Q-learning (MA-QL) algorithm to maximize the energy efficiency, through optimizing the trajectory and transmit power of the UAVs. Finally, simulation results are conducted to verify our analysis and examine the impact of different parameters on the downlink and uplink achievable rates, UAV energy consumption, and system energy efficiency. It is demonstrated that the proposed algorithm achieves much higher energy efficiency than other benchmark schemes.**

***Index Terms*—UAV swarm, energy efficiency, trajectory optimization, multi-agent reinforcement learning, queue theory.**

I. Introduction

Mobile devices and data traffic in edge networks will grow exponentially over the next few years [1]. To meet these demands and provide the holographic coverage for edge networks, it is necessary to develop dynamic, scalable, and self-organized networks. In the last decade, the technology related to autonomous drones, also called Unmanned Aerial Vehicles (UAVs), has been rapidly developed. With small size, high mobility, and low communication overhead, UAVs have emerged as a viable platform to operate in regions where the presence of onboard human pilots is either too risky or unnecessary, from reconnaissance and surveillance tasks for the military [2] to civilian uses such as precision agriculture [3, 4] and logistics [5]. UAVs are also commonly regarded as an effective technique for coverage enhancement in future wireless networks. However, due to the physical limitations of UAVs [6, 7] such as short battery life, it is difficult to rely on a single UAV. Hence, in many applications, two or more UAVs are required to cooperate with each other to complete complex tasks [8].

This work was supported in part by the Natural Science Foundation on Frontier Leading Technology Basic Research Project of Jiangsu under Grant BK20212001, in part by the Project funded by British Council under Grant 913030644, in part by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grants EP/T015985/1 and EP/N007840/1, and in part by the National Natural Science Foundation (NSFC) of China under Grants 62071352 and 92067201. *(Corresponding author: Yongxu Zhu)*

W. Xia is with the Department of Wireless Communication Key Lab of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (email: xiawenchao@njupt.edu.cn).

Y. Zhu, L. De Simone, and T. Dagiuklas are with the Division of Computer Science and Informatics, London South Bank University, London, SE1 0AA, UK (email: {yongxu.zhu, desimol2, tdagiuklas}@lsbu.ac.uk).

K.-K. Wong is with the Department of Electronic and Electrical Engineering, University College London, London, WC1E 7JE, UK (email: kai-kit.wong@ucl.ac.uk).

G. Zheng is with the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Leicestershire, LE11 3TU, UK (email: g.zheng@lboro.ac.uk).

A. Background Work

1) UAV Enabled Wireless Communications: UAVs have been widely used in wireless communications. For example, an analytical framework based on stochastic geometry tools was developed in [9, 10] to evaluate the performance of a three-dimension (3D) UAV network in the presence of interference, which used the Binomial Point Process to model the spatial distribution of the UAVs. In [11], radio interference was analyzed using stochastic geometry theory and a grid-based design of a primary exclusive region for spectrum sharing in a 3D UAV network was presented. In [12–14], a tractable stochastic analysis was employed to characterize the coverage probability of air stations (i.e., UAVs). In [15], a multiple-input multiple-output (MIMO) non-orthogonal multiple access (NOMA) assisted UAV network was investigated, where a stochastic geometry model was established for randomly roaming NOMA users. In [16], a UAV enabled multi-user communication system was studied, where UAV altitude and antenna beamwidth were optimized jointly. Finally, a UAV-enabled communication system was investigated in [17], where a UAV was used to communicate with ground nodes in the presence of multiple jammers.

2) Energy Efficiency of UAV Networks: The energy efficiency of UAV enabled wireless networks has been addressed in various works. In [18], a UAV-enabled wireless communication system with energy harvesting has been investigated, where the total energy consumption of the UAV was minimized while satisfying the minimal data transmission requests of the users. In [19], the energy efficiency was maximized by optimally planning the trajectory of the UAV collecting sensor data from devices scattered around. In [20], an analytical model on the propulsion energy consumption of fixed-wing UAVs has been derived and the UAV's energy efficiency was maximized considering general constraints on its trajectory. Furthermore, the downlink transmission for a

multi-band heterogeneous UAV network was considered in [21], where an efficient coverage radius for the UAVs and an energy-efficient radio resource management scheme were found. In [22], the energy efficiency of a UAV swarm-enabled small cell network was maximized by exploiting the large-scale channel state information at transmitters.

3) Reinforcement Learning (RL) Empowered UAV Networks: RL technique has been widely used in the field of UAVs. In [23], an adaptive federated RL-based jamming attack defense strategy was developed. In [24, 25], a deep RL algorithm was used to compute the optimal trajectory. In [26, 27], two different RL algorithms have been proposed to control the transmit power and to manage interference. In [28], a multi-agent algorithm was investigated to optimize UAVs' sensing tasks and minimize the age of information. The multi-agent RL algorithm in [29] achieved the maximal throughput in a multi-agent downlink network by using predefined UAV trajectory. An autonomous decision-making method for UAV networks was developed in [30] based on deep belief network and Q-learning techniques.

B. Motivation and Contributions

Energy related problems of UAVs have been studied in [18–22, 31, 32], but only a single UAV was applied in [18–21] and energy efficiency was investigated in [32] only for uplink scenarios. Besides, the acceleration energy consumption was not considered in [22, 31]. In brief, the existing literature addresses either the downlink [18, 33] or uplink [34, 35] transmission of UAV-assisted wireless networks, but not both, and also without a realistic energy consumption model. In order to fill this gap, this paper studies a wireless communication system, where the uplink and downlink links are jointly optimized for energy efficiency considering communication related energy and propulsion energy consumption. In the network under study, user terminals (UTs) are served in parallel at the same frequency band. Considering that a UAV working in the full duplex mode for both the downlink and uplink transmission will cause severe self-interference, we introduce two UAVs for the downlink and uplink transmission, respectively. In order to maximize the energy efficiency of the whole network, we propose a joint multi-agent Q-learning (MA-QL) algorithm that simultaneously optimizes the trajectory and transmit power of the UAVs. The contributions of this paper are listed as follows.

- We consider an edge network where two UAVs cooperate with each other and take charge of downlink and uplink transmission, respectively. Two main kinds of interference are considered. In particular, one comes from the UTs working in the transmission mode to the UTs in the receiving mode and the other one is from the emitter UAV to the receiving UTs. The conditional average achievable rates, as well as their lower bounds, for both the uplink and downlink transmission are derived.
- The queue theory is introduced to model downlink and uplink service as two queues and the average hovering time of the UAVs is obtained. Then, we formulate an energy efficiency maximization problem which considers communication related energy and propulsion energy consumption.

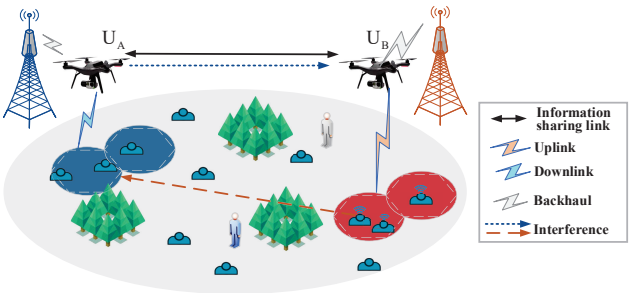


Fig. 1: A system model with two UAVs and a set of UTs.

- We develop a novel MA-QL algorithm to maximize the energy efficiency of the whole network, using a dynamic learning rate and an adaptive ϵ-greedy scheme. In the proposed algorithm, the two UAVs cooperatively take actions via sharing their state information. Simulation results demonstrate that the proposed MA-QL algorithm can achieve better energy efficiency than the zigzag trajectory and random trajectory based approaches.

The rest of the paper is organized as follows. Section II introduces the system model and Section III analyses the downlink and uplink achievable rates. The energy efficiency maximization problem is formulated in Section IV and the MA-QL algorithm is proposed in Section V. The numerical results are presented in Section VI. Finally, we conclude the paper in Section VII.

II. System Model

This paper considers an edge network with two single-antenna UAVs and a set of single-antenna UTs, as shown in Fig. 1. Since no direct communication links are available between the macro base stations (MBSs) and UTs due to far distances, the two UAVs, working as relays, travel between the MBSs to help information transfer. Specifically, one UAV, called $\mathrm{U_A}$, is responsible for downlink transmission while the other UAV, called $\mathrm{U_B}$, is responsible for data collection from the UTs in the uplink. UAVs $\mathrm{U_A}$ and $\mathrm{U_B}$ are assumed to be elevated at fixed altitude H_A and H_B, respectively. The UTs are geographically distributed in a certain destination area Ω according to a homogeneous Poisson Point Process (PPP) with spatial density λ. Given that the active probability of each UT is $\mathrm{Pr^{Ac}}$, then the spatial density of the active UTs is $\tilde{\lambda} = \mathrm{Pr^{Ac}}\lambda$.

The two UAVs, $\mathrm{U_A}$ and $\mathrm{U_B}$, have different footprints with radii r_A and r_B, respectively. For UAV $\mathrm{U_A}$, the destination area Ω can be approximately divided into $K_\mathrm{A} = \frac{||\Omega||}{\pi r_\mathrm{A}^2}$ sub-areas that each can be entirely covered by the footprint of UAV $\mathrm{U_A}$, where $||\Omega||$ is the area size. The index set of the sub-areas is denoted as $\mathcal{K}_\mathrm{A} = \{1, 2, \ldots, K_\mathrm{A}\}$. Similarly, the number of the sub-areas for UAV $\mathrm{U_B}$ is approximately equal to $K_\mathrm{B} = \frac{||\Omega||}{\pi r_\mathrm{B}^2}$ and the index set of the sub-areas is denoted as $\mathcal{K}_\mathrm{B} = \{1, 2, \ldots, K_\mathrm{B}\}$.

Without loss of generality, we assume the two UAVs $\mathrm{U_A}$ and $\mathrm{U_B}$ serve the UTs by covering the sub-areas one by one. Once the two UAVs reach the next sub-area, they stop to hover over the sub-area and start fulfilling requests from the UTs.

A. Channel Model

The communication links between the UAVs and UTs can be either line-of-sight (LoS) or non-LoS (NLoS) with different probabilities of occurrences. The probability of having a LoS link between UAV $\mathrm{U}_i, i \in \{\mathrm{A}, \mathrm{B}\}$ and UT u is formulated as [29]

$$\Pr^{\mathrm{L}}(x_{i,u}) = \frac{1}{1 + \kappa_1 \exp(-\kappa_2 \tan^{-1}(\frac{H_i}{x_{i,u}}) - \kappa_1)}, \quad (1)$$

where κ_1 and κ_2 are two constants that depend on the environment, and $x_{i,u}$ is the projection distance between UAV U_i and UT u. Then, the probability of having a NLoS link is $\Pr^{\mathrm{N}}(x_{i,u}) = 1 - \Pr^{\mathrm{L}}(x_{i,u})$. Denote the LoS and NLoS path loss as $\mathrm{PL}^{\mathrm{L}}_{i,u} = d^{-\alpha^{\mathrm{L}}}_{i,u}$ and $\mathrm{PL}^{\mathrm{N}}_{i,u} = d^{-\alpha^{\mathrm{N}}}_{i,u}$, respectively, where $d_{i,u} = \sqrt{x^2_{i,u} + H^2_i}$ is the distance between UT u and U_i and α^{L} and α^{N} are the path loss exponents for LOS and NLOS links, respectively. Thus, the average path loss between UAV U_i and UT u is expressed as

$$\overline{\mathrm{PL}}_{i,u} = \Pr^{\mathrm{L}}(x_{i,u})\mathrm{PL}^{\mathrm{L}}_{i,u} + \Pr^{\mathrm{N}}(x_{i,u})\mathrm{PL}^{\mathrm{L}}_{i,u}. \quad (2)$$

Besides, we assume the communication links between the UTs are NLoS whereas the communication links between the UAVs are LoS.

The Nakagami-m distribution is a universal model suitable for various conditions, which can be used to model not only the Rician fading in the LoS scenarios, but also the Rayleigh fading in the NLoS scenarios [36]. Thus, here we assume all transmission links experience independent Nakagami-m distribution and the small-scale fading gain follows a Gamma distribution $\Gamma(x, y)$.

B. Downlink Transmission

We first consider the downlink transmission from UAV U_A to a typical UT u. Define P_A and P_u as the transmit power of UAV U_A and UT u, respectively. Then, the received signal-to-interference-plus-noise ratio (SINR) from UAV U_A to UT u is given as [37]

$$\gamma_\mathrm{A} = \frac{P_\mathrm{A}\beta_0 h_{\mathrm{A},u}\overline{\mathrm{PL}}_{\mathrm{A},u}}{\delta \mathrm{Int}^{\mathrm{dl}} + \sigma^2}, \quad (3)$$

where $h_{\mathrm{A},u} \sim \Gamma(\phi_1, 1/\phi_1)$ denotes the small-scale fading gain, ϕ_1 is the fading parameter, σ^2 is the additive noise power, β_0 is the path loss at the reference distance $d_0 = 1$m, and δ is an indicator function defined as

$$\delta = \begin{cases} 1, \text{if } \mathrm{U}_\mathrm{A} \text{ and } \mathrm{U}_\mathrm{B} \text{ are hovering at the same time,} \\ 0, \text{else.} \end{cases} \quad (4)$$

Note that when U_A transmits downlink signals to UT u, the active UTs that send their uplink signals at the same time will cause inter-cell interference to UT u. We collect these active UTs that send uplink signals into a set Φ_B. Then, the inter-cell interference $\mathrm{Int}^{\mathrm{dl}}$ from Φ_B is given by

$$\mathrm{Int}^{\mathrm{dl}} = P_u g_u \beta_0 \bar{d}^{-\alpha^{\mathrm{N}}}_{u,\Phi_\mathrm{B}}, \quad (5)$$

where $g_u \sim \Gamma(1, 1)$ is the small-scale fading gain and $\bar{d}_{u,\Phi_\mathrm{B}}$ is the average distance between the typical UT u and the UTs in Φ_B, which is computed as

$$\bar{d}_{u,\Phi_\mathrm{B}} = \frac{\sum_{u' \in \Phi_\mathrm{B}} d_{u,u'}}{|\Phi_\mathrm{B}|}, \quad (6)$$

where $d_{u,u'}$ is the distance between the two UTs u and u' and $|\Phi_\mathrm{B}|$ is the cardinality of Φ_B. Then, the achievable rate is given as $R_\mathrm{A} = \log_2(1 + \gamma_\mathrm{A})$.

C. Uplink Transmission

Without loss of generality, we choose a typical UT u in Φ_B as an example and the received SINR of UT u at UAV U_B is given as

$$\gamma_\mathrm{B} = \frac{P_u \beta_0 h_{\mathrm{B},u}\overline{\mathrm{PL}}_{\mathrm{B},u}}{\delta \mathrm{Int}^{\mathrm{ul}} + \sigma^2}, \quad (7)$$

where $h_{\mathrm{B},u} \sim \Gamma(\phi_2, 1/\phi_2)$ denotes the small-scale fading gain, ϕ_2 is the fading parameter, and the inter-cell interference $\mathrm{Int}^{\mathrm{ul}}$ from U_A is given as

$$\mathrm{Int}^{\mathrm{ul}} = P_\mathrm{A} g_{\mathrm{A,B}} |d_{\mathrm{A,B}}|^{-\alpha^{\mathrm{L}}}, \quad (8)$$

where $g_{\mathrm{A,B}} \sim \Gamma(\phi_3, 1/\phi_3)$ denotes the small-scale fading gain, ϕ_3 is the fading parameter, and $d_{\mathrm{A,B}}$ is the distance between U_A and U_B. Then, the achievable rate is given as $R_\mathrm{B} = \log_2(1 + \gamma_\mathrm{B})$.

D. Queue Model

The two UAVs serve their respective sub-areas one by one and before entering a target sub-area, we assume that there have been $L^0_i, i \in \{\mathrm{A}, \mathrm{B}\}$, UTs waiting for service in the target sub-area. During the hovering time on the target sub-area, new service requests are generated from the UTs and we model the new arrivals as a queue. The arrival rate follows a Poisson distribution with parameter ν_i and the service time follows an exponential distribution with parameter μ_i. Given that the uplink/downlink transmission can be considered as a finite $M/M/1$ queue with an initial length L^0_i, the parameters μ_i of the mean service time is computed as

$$\mu_i = \frac{\bar{R}_i}{\bar{Q}_i}, \quad (9)$$

where $\bar{Q}_i$ is the amount of the approximate data requested and $\bar{R}_\mathrm{A} = \mathbb{E}[R_\mathrm{A}]$ and $\bar{R}_\mathrm{B} = \mathbb{E}[R_\mathrm{B}]$ are the conditional average downlink and uplink achievable rates, respectively, for a typical UT, which are specified in the following section.

According to [38], the average hovering time for U_i is

$$t^{\mathrm{Hov}}_i = \frac{1}{\mu_i - \nu_i} + \frac{L^0_i}{\mu_i}, i \in \{\mathrm{A}, \mathrm{B}\}. \quad (10)$$

Moreover, the total length of the queue needs to be less than the total number of the active UTs in the target area, i.e.,

$$\frac{\nu_i}{\mu_i - \nu_i} + L^0_i \leq \rho_i, \quad (11)$$

where $\rho_i = \tilde{\lambda}\pi r^2_i, i \in \{\mathrm{A}, \mathrm{B}\}$. Furthermore, we have $\nu_i < \mu_i$ in order to keep the two queues stable.

III. Performance Analysis

In this section, we analyze the average achievable rates in both the downlink and uplink transmission.

A. Downlink Achievable Rate

***Theorem** 1:* Given the projection distance x between $\mathrm{U_A}$ and a typical UT u and $d_{\mathrm{A}} = \sqrt{x^2 + H_{\mathrm{A}}^2}$, then the conditional average downlink achievable rate of UT u served by $\mathrm{U_A}$ can be expressed as

$$\bar{R}_{\mathrm{A}} = \frac{2\pi}{\ln 2} \int_0^{r_{\mathrm{A}}} \int_0^{\infty} \frac{\Pi_{\mathrm{A}}^{\mathrm{N}}(\chi, x)\mathrm{Pr}^{\mathrm{N}} + \Pi_{\mathrm{A}}^{\mathrm{L}}(\chi, x)\mathrm{Pr}^{\mathrm{L}}}{1+\chi} x \mathrm{d}\chi \mathrm{d}x, \tag{12}$$

where $\xi = \frac{\chi}{P_{\mathrm{A}}\beta_0}$,

$$\Pi_{\mathrm{A}}^{\mathrm{L}}(\chi, x) = \sum_{j=1}^{\phi_1} (-1)^{j+1} \binom{\phi_1}{j} e^{-\xi d_{\mathrm{A}}^{\alpha^{\mathrm{L}}}(\sigma^2 + \delta\Psi_1)}, \tag{13}$$

$$\Pi_{\mathrm{A}}^{\mathrm{N}}(\chi, x) = e^{-\xi d_{\mathrm{A}}^{\alpha^{\mathrm{N}}}(\sigma^2 + \delta\Psi_1)}, \tag{14}$$

$$\Psi_1 = P_u\beta_0 \left[\frac{\int_0^{2\pi}\int_0^{r_{\mathrm{A}}} \Xi(\Delta\bar{D}, x, \theta) x \mathrm{d}x \mathrm{d}\theta}{\pi r_{\mathrm{A}}^2} \right]^{-\alpha^{\mathrm{N}}}, \tag{15}$$

$$\Xi(z_1, z_2, \vartheta) = ({z_1}^2 + {z_2}^2 - 2z_1 z_2 \cos\vartheta)^{1/2}, \tag{16}$$

and

$$\Delta\bar{D} = \frac{\int_0^{2\pi}\int_0^{r_{\mathrm{B}}} \Xi(d_{\mathrm{A,B}}, x', \theta_2) x' \mathrm{d}x' \mathrm{d}\theta_2}{\pi r_{\mathrm{B}}^2}. \tag{17}$$

Proof: See Appendix A.

***Theorem** 2:* A tractable lower bound for the conditional average achievable rate $\bar{R}_{\mathrm{A}}$ of the typical UT u served by $\mathrm{U_A}$ in the downlink can be computed as

$$\bar{R}_{\mathrm{A}}^{\mathrm{lb}} = \frac{1}{\pi r_{\mathrm{A}}^2} \int_0^{2\pi} \int_0^{r_{\mathrm{A}}} \log_2(1 + \Delta_{\mathrm{A}}) x \mathrm{d}x \mathrm{d}\theta, \tag{18}$$

where Δ_{A} is expressed as

$$\Delta_{\mathrm{A}} = \frac{P_{\mathrm{A}}\beta_0 \left(\mathrm{Pr}^{\mathrm{L}} d_{\mathrm{A}}^{-\alpha^{\mathrm{L}}} + \mathrm{Pr}^{\mathrm{N}} d_{\mathrm{A}}^{-\alpha^{\mathrm{N}}}\right)}{\delta P_u\beta_0 \left[\frac{\int_0^{2\pi}\int_0^{r_{\mathrm{A}}} \Xi(\Delta\bar{D}, x, \theta) x \mathrm{d}x \mathrm{d}\theta}{\pi r_{\mathrm{A}}^2} \right]^{-\alpha^{\mathrm{N}}} + \sigma^2}. \tag{19}$$

Proof: See Appendix B.

B. Uplink Achievable Rate

***Theorem** 3:* Given the projection distance x' between $\mathrm{U_B}$ and a typical UT u' and $d_{\mathrm{B}} = \sqrt{(x')^2 + H_{\mathrm{B}}^2}$, then the conditional average uplink achievable rate of UT u' served by $\mathrm{U_B}$ can be expressed as

$$\bar{R}_{\mathrm{B}} = \frac{2\pi}{\ln 2} \int_0^{r_{\mathrm{B}}} \int_0^{\infty} \frac{\Pi_{\mathrm{B}}^{\mathrm{L}}(\chi, x')\mathrm{Pr}^{\mathrm{L}} + \Pi_{\mathrm{B}}^{\mathrm{N}}(\chi, x')\mathrm{Pr}^{\mathrm{N}}}{1+\chi} x' \mathrm{d}\chi \mathrm{d}x', \tag{20}$$

where $\hat{\xi} = \frac{\chi}{P_u\beta_0}$,

$$\Pi_{\mathrm{B}}^{\mathrm{L}}(\chi, x') \approx \sum_{j=1}^{\phi_2} (-1)^{j+1} \binom{\phi_2}{j} e^{-\hat{\xi} d_{\mathrm{B}}^{\alpha^{\mathrm{L}}}(\sigma^2 + \delta\Psi_2)}, \tag{21}$$

and

$$\Pi_{\mathrm{B}}^{\mathrm{N}}(\chi, x') = e^{-\hat{\xi}\chi d_{\mathrm{B}}^{\alpha^{\mathrm{N}}}(\sigma^2 + \delta\Psi_2)}, \tag{22}$$

with $\Psi_2 = P_{\mathrm{A}}\beta_0 d_{\mathrm{A,B}}^{-\alpha^{\mathrm{L}}}$.

Proof: See Appendix C.

***Theorem** 4:* A tractable lower bound for the conditional average achievable rate $\bar{R}_{\mathrm{B}}$ of the typical UT u' served by $\mathrm{U_B}$ in the uplink can be computed as

$$\bar{R}_{\mathrm{B}}^{\mathrm{lb}} = \frac{1}{\pi r_{\mathrm{B}}^2} \int_0^{2\pi} \int_0^{r_{\mathrm{B}}} \log_2(1 + \Delta_{\mathrm{B}}) x' \mathrm{d}x' \mathrm{d}\theta_2, \tag{23}$$

with

$$\Delta_{\mathrm{B}} = \frac{P_u\beta_0 \left(\mathrm{Pr}^{\mathrm{L}} d_{\mathrm{B}}^{-\alpha_{\mathrm{L}}} + \mathrm{Pr}^{\mathrm{N}} d_{\mathrm{B}}^{-\alpha_{\mathrm{N}}}\right)}{\delta P_{\mathrm{A}}\beta_0 d_{\mathrm{A,B}}^{-\alpha_{\mathrm{L}}} + \sigma^2}. \tag{24}$$

Proof: See Appendix D.

IV. Problem Formulation

In this section, we analyze the energy consumption of the UAVs, which includes two main components, i.e., communication related energy and propulsion energy consumption. For a rotary-wing UAV, its speed can be computed as

$$v(t) = \frac{6D}{t^2}(\tau_{k,k+1} + \tau_{k,k+1}^2), \tag{25}$$

where $\tau_{k,k+1}$ is the transition time from sub-area k to the next, D is the space travelled, and the mechanical energy consumption is given as

$$E_i^{\mathrm{Mec}}(t) = \sum_{k \in \mathcal{K}_i \setminus \{K_i\}} \left[\int_0^{\tau_{k,k+1}} P^{\mathrm{Total}}(t) dt + E_i^{\mathrm{Kin}} \right], \tag{26}$$

where E_i^{Kin} is the change in kinetic energy and $P^{\mathrm{Total}}(t) = P^{\mathrm{BP}}(t) + P^{\mathrm{IN}}(t) + P^{\mathrm{PD}}(t)$ [39]. Specifically, the blade profile power $P^{\mathrm{BP}}(t) = c_1(1 + c_2 v^2(t))$ is a unique term for rotary-wing UAV [40], which is required to overcome the profile drag due to the rotation of blades. The induced power, denoted as

$$P^{\mathrm{IN}}(t) = c_3 \sqrt{1 + \frac{\varpi^2(t)}{\psi^2}} \left(\sqrt{1 + \frac{\varpi^2(t)}{\psi^2} + \frac{v^4(t)}{c_4^2}} - \frac{v^2(t)}{c_4} \right)^{1/2}, \tag{27}$$

is required to overcome the induced drag developed during the creation of the lift force to maintain the aircraft airborne, where ψ is the gravitational acceleration and $\varpi(t)$ is the UAV acceleration.

The parasite power $P^{\mathrm{Parasite}} = c_5 v^3(t)$ is the component required to overcome the parasite friction drag due to the movement of the aircraft in the air. E_i^{Kin} is the change in kinetic energy, i.e.,

$$E_i^{\mathrm{Kin}} = \frac{1}{2} M(|v_i(\tau_{k,k+1})|^2 - |v_i(0)|^2), \tag{28}$$

where $v_i(0)$ and $v_i(\tau_{k,k+1})$ are the initial and final speeds of U_i, respectively, when transiting from subarea k to subarea $k+1$, and M is the mass of U_i. Moreover, $c_i \in [1, 2, ..., 5]$ are the modelling parameters that depend on the UAV weight, air density, and rotor disc area, as specified in [39]. In particular,

we can define c_1 and c_5 as parameters linked to the rotor features and air density, c_2 as the parameter related to the blade angular velocity, while c_4 is connected to the rotor velocity, and c_3 is the parameter taking into account the UAV weight. Note that the UAV acceleration and speed functions have been computed using an interpolation technique as in [41]. The polynomial function used is cubic as we are in a system with constant acceleration/deceleration. The parasite power exists only when the UAV has a nonzero flying speed. Both the blade profile power and the parasite power increase with the growth of the aircraft speed v while the induced power decreases as v increases.

The communication-related energy for the UAVs is computed as [42]

$$E^{\mathrm{Com}} = T_{\mathrm{A}}^{\mathrm{Hov}} P_{\mathrm{A}} + T_{\mathrm{B}}^{\mathrm{Hov}} \rho_{\mathrm{B}} P_u, \tag{29}$$

where $T_i^{\mathrm{Hov}} = K_i t_i^{\mathrm{Hov}}$ is the total hovering time.

For the static speed $v(t) \to 0$, the power consumption corresponding to the hovering UAV at the fixed location is asymptotically derived as

$$E_i^{\mathrm{Hov}} = T_i^{\mathrm{Hov}}(c_1 + c_3). \tag{30}$$

Define the energy efficiency of the whole system as

$$\mathtt{EE} = \frac{\sum_{i\in\{\mathrm{A,B}\}} \omega_i K_i \rho_i \bar{R}_i}{\sum_{i\in\{\mathrm{A,B}\}} E_i^{\mathrm{Mec}} + \sum_{i\in\{\mathrm{A,B}\}} E_i^{\mathrm{Hov}} + \hat{\omega} E^{\mathrm{Com}}}, \tag{31}$$

where ω_i and $\hat{\omega}$ are weighting parameters. In this work, we aim to maximize the energy efficiency of the whole system, which is formulated as

$$\mathcal{P}1: \max_{\Theta(t)} \mathbb{E}\left[\mathtt{EE}\right] \tag{32}$$

$$\text{s.t. } \mathrm{C1}: P_{\mathrm{A}} \le P_{\mathrm{A,max}}, \tag{33}$$

$$\mathrm{C2}: \delta \in \{0,1\}, \tag{34}$$

$$\mathrm{C3}: v(t) \le V_{\max}, \tag{35}$$

where $V_{\max}$ is the maximum value of $v(t)$, $P_{\mathrm{A,max}}$ is the transmit power budget of $\mathrm{U_A}$, and Θ is a triple denoted as $\Theta = \langle \mathbf{c}_{\mathrm{A}}, \mathbf{c}_{\mathrm{B}}, P_{\mathrm{A}} \rangle$, with $\mathbf{c}_{\mathrm{A}}$ and $\mathbf{c}_{\mathrm{B}}$ being the subarea sequence of the trajectories of $\mathrm{U_A}$ and $\mathrm{U_B}$. Problem $\mathcal{P}1$ is a complicated optimization problem, which cannot be solved in polynomial time. RL is an efficient way to solve the model-free and complex problems by taking appropriate actions to maximize the reward in a certain environment, which sheds lights on problem $\mathcal{P}1$. In the following, we will present the framework of the multi-agent RL method for solving the multi-UAV cooperative resource allocation problem.

V. Proposed Algorithm

According to [43], we model the energy efficiency maximization problem in problem $\mathcal{P}1$ as a Markov game, which can be regarded as a multi-agent extension of the Markov decision process. In the Markov game, the two UAVs, $\mathrm{U_A}$ and $\mathrm{U_B}$, are considered as two agents. The Markov game is composed of a state space $\mathcal{S}$, an action space $\mathcal{A}$, and a reward space $\mathbf{\Phi}$. To achieve better performance, we assume that the two UAVs can interact with each other and exchange some auxiliary information. However, only a few bits are allowed to be exchanged since there is no wired link available between the UAVs.

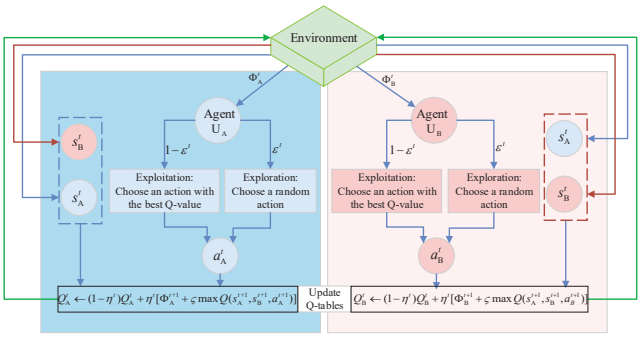


Fig. 2: The proposed MA-QL algorithm framework.

Agent: We assume that the two agents, $\mathrm{U_A}$ and $\mathrm{U_B}$, exchange information regarding their current states, as shown in Fig. 2. With the state information, the two agents take actions according to their policies, i.e., π_{A} and π_{B}, and receive their respective rewards, i.e., Φ_{A}^t and Φ_{B}^t.

State: Define the state space of agent U_i as $\mathcal{S}_i \triangleq (k_i, \gamma_i)$, where the first value represents the current position (i.e., sub-area k_i) of agent U_i and the second value represents the achievable SINR value. Since γ_i is a continuous variable, the number of possible states can be huge. To address this challenge, we divide the value range $[\gamma_{i,\min}, \gamma_{i,\max}]$ of γ_i into N_i segments $[\gamma_{i,n}, \gamma_{i,n+1}]_{n=0}^{N_i-1}$. Thus, the state space is reduced to $\mathcal{S}_i \triangleq (k_i, n_i)$, where n_i indicates that γ_i belongs to the n-th segment i.e., $\gamma_i \in [\gamma_{i,n-1}, \gamma_{i,n}]$.

Action: The action set $\mathcal{A}_{\mathrm{A}} \triangleq (m_{\mathrm{A}}, P_{\mathrm{A}})$ of $\mathrm{U_A}$ contains the movement and the transmit power actions, where m_A represents the movement direction on a 2D surface as $m_A \in \mathcal{M} = \{0,1,2,3,4,5,6\}$, with the value 0 indicating the hovering action at the same position, as shown in Fig. 3. Similar to γ_i, the transmit power P_{A} is also a continuous variable and thus the value range $[0, P_{\mathrm{A,max}}]$ of P_{A} is divided into Z_{A} segments $[P_{\mathrm{A},z}, P_{\mathrm{A},z+1}]_{z=0}^{Z_A-1}$. Consequently, the action set of agent $\mathrm{U_A}$ is reduced to $\mathcal{A}_{\mathrm{A}} \triangleq (m_{\mathrm{A}}, z_{\mathrm{A}})$. Different from agent $\mathrm{U_A}$, the action space of agent $\mathrm{U_B}$ contains only the movement action, i.e., $\mathcal{A}_{\mathrm{B}} \triangleq (m_{\mathrm{B}})$, because P_u is managed by the UTs.

Reward: The reward function of agent U_i at each iteration t is defined as

$$\Phi_i(\varrho_i^t) = \frac{\rho_i \bar{R}_i [1 + (K_i \rho_i - G_i^t)]^{-1}}{E_i^{\mathrm{Mec}} + E_i^{\mathrm{Hov}} + E^{\mathrm{Com}}}, \tag{36}$$

where $\varrho_i^t = (s_i^t, s_{-i}^t, a_i^t)$ with $-i = \{\mathrm{A,B}\} \setminus \{i\}$ and G_i^t is the number of users already served by agent U_i in the previous iterations, which is computed as

$$G_i^t = \zeta_i^t (L_i^0 + \frac{\nu_i}{\mu_i - \nu_i}), \tag{37}$$

with $\zeta_i^t \le K_i$ as the number of sub-areas already covered by U_i previously. The reward function in (36) enables three objectives: maximizing the average achievable rate for the downlink/uplink transmission, maximizing the coverage area for the active UTs, and minimizing the energy consumption. The term $K_i \rho_i^t - G_i^t$ in (36) can be considered as the effective incremental coverage, which adds a penalty to the actual value.

In summary, $\bar{R}_i$ gives the gain while the denominator is the cost in terms of energy consumption. The agents try to select the actions such that the sum of the discounted rewards is maximized in the future [44]. Maximizing the cumulative reward is approximately equivalent to maximizing the energy efficiency.

Probability of action selection: In a classic Q-learning approach, the ϵ-greedy scheme is usually used to strike a balance between exploration and exploitation such that the agents reinforce the actions performed well in the past but also explore new actions that might return higher rewards in the future. Instead of using a static ϵ value, in this paper we employ a dynamic ϵ, i.e.,

$$a_i^t = \begin{cases} \arg\max_{a_i} Q_i(\varrho_i^t), & 1-\epsilon^t, \\ \text{random}, & \epsilon^t, \end{cases} \tag{38}$$

where $Q_i(\varrho_i^t)$ is the Q-value in the Q-table of U_i and ϵ^t is the probability of choosing a random action in iteration t, whose update rule is

$$\epsilon^t = \epsilon_{\max} - \sum_{t=1}^{t/\Delta t} \frac{(\epsilon_{\max} - \epsilon_{\min})\Delta t}{T}, \tag{39}$$

where $\epsilon_{\min}$ and $\epsilon_{\max}$ are the minimum and maximum values of ϵ, respectively, Δt is a constant interval, and T is the expected number of training iterations. In (39), ϵ decreases every Δt iterations until reaching the minimum value $\epsilon_{\min}$. This initial setting $\epsilon(1) = \epsilon_{\max}$ allows the agents first to explore the environment and enrich the Q-tables. A smaller ϵ will lead to the agents taking more efficient actions and achieving better Q-values. For each action taken, the agents receive a reward and new values are calculated for each state-action pair. The algorithm uses an iterative updating and correction process based on the new information, and the Q-tables are updated simultaneously [45]. The Q-value update function for U_i is computed as

$$Q_i(\varrho_i^t) \leftarrow (1-\eta^t) Q_i(\varrho_i^t) + \eta^t [\Phi_i^{t+1} + \varsigma \max_{a_i^{t+1} \in \mathcal{A}_i} Q_i(\varrho_i^{t+1})], \tag{40}$$

where $\varsigma \in (0, 1]$ is a discount factor. An episode of the algorithm ends when the state s_{B}^t or s_{A}^t is a final state or a state of absorption. The final state is defined as the state in which the agents have served all the users.

We aim to find an optimal policy $\pi^\star : \mathcal{S}_i \rightarrow \mathcal{A}_i$ for the agents to maximize the expected long-term reward function in the system. Accordingly, we define value function $V_i^\pi : s_i \rightarrow \Phi_i$ that represents the expected value obtained by the following policy π_i of each state $s_i^t \in \mathcal{S}_i$. The optimal action at each state can be found through the optimal value function expressed by

$$V_i^\star(s_i^t) = \max_{a_i^t} \{\mathbb{E}[\Phi_i^{t+1} + \varsigma V_i^\star(s_i^{t+1})]\}. \tag{41}$$

Note that $V_i^\star(s_i^t) = \max_{a_i^t}\{Q_i^{\pi^\star}(\varrho_i^t)\}$ and $Q_i^{\pi^\star}(\varrho_i^t) = \mathbb{E}[\Phi_i^{t+1} + \varsigma V_i^\star(s_i^{t+1})]$ is the optimal Q-function for all state$_{\mathrm{A}}$-state$_{\mathrm{B}}$-action$_i$ triplets.

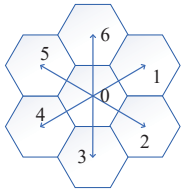


Fig. 3: Action directions.

The proposed MA-QL algorithm is presented in Algorithm 1. In the proposed MA-QL algorithm, It will return as outputs the trajectory and the list of the transmit power values used, for each UAV. For each iteration, the algorithm updates the dynamic parameters and the Q-tables, respectively. It can be seen that in each iteration, the total iteration reward is computed as the sum of each Φ_t^i, until both the UAVs reach the final state.

In the proposed algorithm, an adaptive update rule of the learning rate is given as

$$\eta^t = \left[\eta_{\max} + \sum_{t=1}^{t/\Delta t} \frac{(\eta_{\max}/\eta_{\min})\Delta t}{T}\right]^{-1}, \eta_{\min} \leq \eta(t) \leq \eta_{\max}, \tag{42}$$

where η decreases every Δt iterations with the factor $\frac{(\eta_{\max}/\eta_{\min})\Delta t}{T}$. The discrete curve that is generated has a very sharp initial inclination, which then decreases. The dynamic learning rate speeds up the training process while guarantees the convergence of the algorithm, as proved in [46]. Then the dynamic learning rate allows us to initially consider most of the episodes experienced previously to test different paths and enrich the Q-table values. At last, as the training iterations pass, it focuses on the last episodes experienced, defined by the highest Q-values in the tables.

Algorithm 1 The proposed MA-QL algorithm.

1: Initialize the Q-table, τ, $\eta(1)$, $\epsilon(1)$, $s_i(0), i \in \{\mathrm{A}, \mathrm{B}\}$, and $t = 1$.
2: **for** each episode t **do**
3: **if** $t \bmod \Delta t = 0$ **then**
4: Update η^t and ϵ^t according to (39) and (42), respectively.
5: **end if**
6: **if** rand($\bullet$) $< \epsilon$ **then**
7: Randomly select action a_i^t.
8: **else**
9: Select action a_i^t = $\arg\max_{a_i} Q_i^t$.
10: **end if**
11: Execute action a_i^t and observe the subsequent state s_i^t.
12: Receive an immediate reward Φ_i^t and update Q-table according to (40).
13: $t = t + 1$.
14: **end for**

Computational Complexity: The complexity of the algorithm has two main contributors, the Q-value updates of agents U_{A} and U_{B}. Given the complexity of calculating the

TABLE I. PARAMETER VALUES.

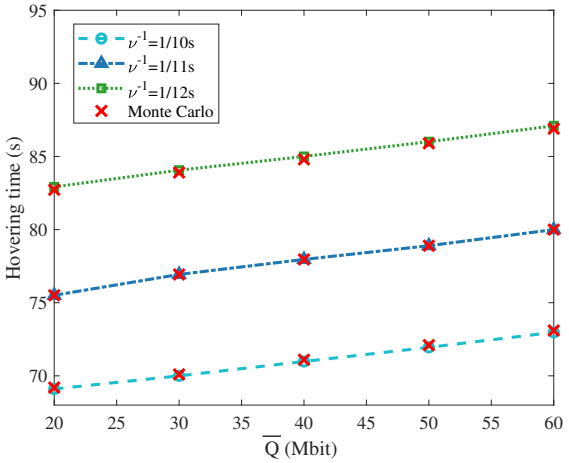
Parameters	Values
mmWave bandwidth (BW)	3 GHz
Noise figure (Nf)	10 dB
Carrier frequency (f_c)	1 GHz
Maximum of $\mathrm{U_A}$ transmit power ($P_{\mathrm{A,max}}$)	46 dBm
Maximum of UT transmit power ($P_{u,\max}$)	30 dBm
Urban environment parameters (κ_1, κ_2)	9.6, 0.28
$\mathrm{U_A}$,$\mathrm{U_B}$ footprint radius (r_{A}, r_{B})	90 m, 90 m
$\mathrm{U_A}$,$\mathrm{U_B}$ altitude (H_{A}, H_{B})	70 m, 80 m
UT density (λ)	250 UTs/km^2
UT active probability ($\Pr^{\mathrm{Ac}}$)	0.8
Path loss exponent ($\alpha^{\mathrm{L}}, \alpha^{\mathrm{N}}$)	2, 3 [47]
Discount factor (ς)	0.9
Maximum/minimum of learning rate ($\alpha_{\max}/\alpha_{\min}$)	1, 0.1
Maximum/minimum of ϵ ($\epsilon_{\max}/\epsilon_{\min}$)	0.95, 0.4
Number of training iterations/episodes (T)	250
Update interval (Δt)	10
Propulsion modelling parameters (c_i)	$c_1 = 580.65$, $c_2 = 7.5e^{-5}$, $c_3 = 3$, $c_4 = 2$, $c_5 = 7.3 \times 10^{-3}$ [39]

Q-value once is $\mathcal{O}(1)$ and the complexity of the Q-value update in each iteration is $\mathcal{O}(|\mathcal{S}_i||\mathcal{A}_i|)$ for agent U_i. More specifically, the sizes of the state space and action space of $\mathrm{U_A}$ are $|\mathcal{S}_{\mathrm{A}}| = K_{\mathrm{A}}N_{\mathrm{A}}$ and $|\mathcal{A}_{\mathrm{A}}| = |m_{\mathrm{A}}|Z_{\mathrm{A}}$, respectively. Similarly, the sizes of the state space and action space of $\mathrm{U_B}$ are $|\mathcal{S}_{\mathrm{B}}| = K_{\mathrm{B}}N_{\mathrm{B}}$ and $\mathcal{A}_{\mathrm{B}} = |m_{\mathrm{B}}|$, respectively. Therefore, the overall computational complexity can be denoted as $\mathcal{O}\Big(T(K_{\mathrm{A}}N_{\mathrm{A}}|m_{\mathrm{A}}|Z_{\mathrm{A}}+K_{\mathrm{B}}N_{\mathrm{B}}|m_{\mathrm{B}}|)\Big)$, where T is the number of all training iterations.

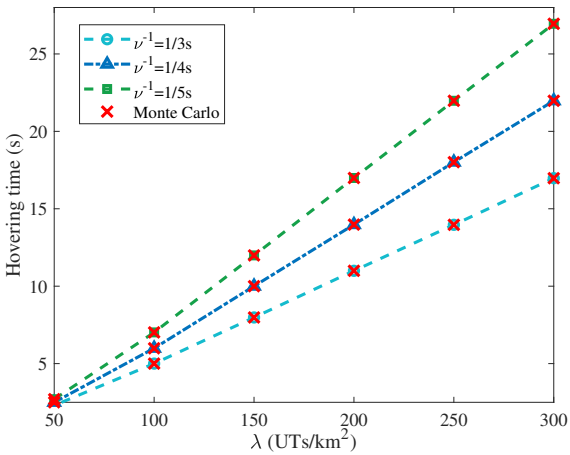
VI. Simulations Results

In this section, we evaluate the performance of the proposed algorithm with numerical results. We consider an area of 1 km^2 and the horizontal locations of the UAVs are restricted in the area. We assume that the noise power is $\sigma^2 = -174 + 10\log_{10}(\mathrm{BW}) + \mathrm{Nf}$ dBm, where BW is the bandwidth and Nf is the noise figure. More parameters are listed in Table I.

Fig. 4 shows the results of the UAV hovering time with respect to the requested data size and the UT density. The solid curves are obtained from (10) and Monte Carlo results are averaged over 5,000 independent trials, with $r_{\mathrm{A}} = r_{\mathrm{B}} = 90$ m, $H_{\mathrm{A}} = H_{\mathrm{B}} = 80$ m, $\lambda = 250$ UTs/km^2, $\mu^{-1} = 0.5$ s, and $\Pr^{\mathrm{Ac}} = 0.8$. Fig. 4(a) illustrates that the hovering time increases as the requested bits of the UTs increase. We can also observe that a larger arrival rate ν leads to higher hovering time due to more service requests from the UTs. Fig. 4(b) shows a rising tendency of the hovering time with the increasing density. Notice that all our analytical results match very well with those via Monte Carlo simulation.



(a) UAV hovering time versus the requested data size.



(b) UAV hovering time versus the UT density.

Fig. 4: UAV hovering time versus the requested data size and the UT density.

A. Effects of Dynamic Parameters

Here we analyze the effects of the dynamic parameters on processing time compared with other four baseline approaches: 1) one with fixed η and dynamic ϵ, 2) one with linearly dynamic η and ϵ, 3) one with dynamic η and fixed ϵ, and 4) one with fixed η and ϵ.

Fig. 5 shows the algorithm processing time of the five parameter pairs against the footprint radius r of the two UAVs with $r = r_{\mathrm{A}} = r_{\mathrm{B}}$. All the dynamic parameters have $\alpha_{\min} = 0.1$ and $\alpha_{\max} = 1$, $\epsilon_{\min} = 0.4$ and $\epsilon_{\max} = 0.95$. Furthermore, the discount factor is fixed as $\gamma = 0.9$. Results illustrate that the proposed approach significantly decreases the algorithm processing time. Specifically, compared with the static and the only-η dynamic approaches (green and purple bars), the proposed approach decreases the time with peaks of 70% less, while compared to the linearly dynamic approach (orange bars) the time decrease reaches up to 25%, and compared to the only-ϵ dynamic approach (red bars) the time is slightly lower. More importantly, the minimum processing time is obtained with $r_{\mathrm{A}} = r_{\mathrm{B}} = 90$ m for all the approaches. At this value, it has reached the trade-off between the computational time for the flying movement and the computational time for interference. Indeed, the number of sub-flights done by the UAVs from one group of the UTs to another increases with a smaller radius, and statistically, the number of times the UAVs serve simultaneously two

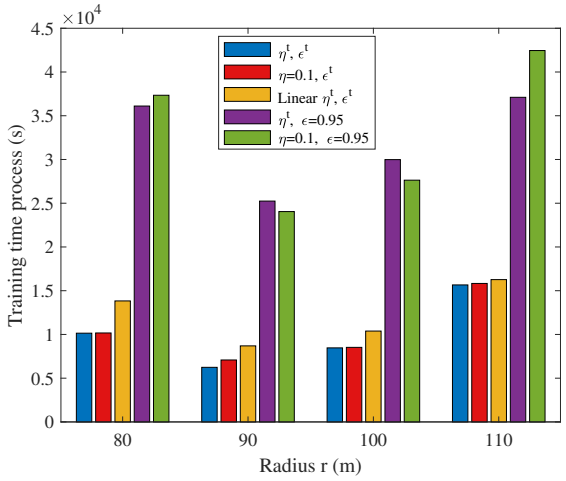


Fig. 5: Training time against the footprint radius ($r = r_{\mathrm{A}} = r_{\mathrm{B}}$) of the UAVs.

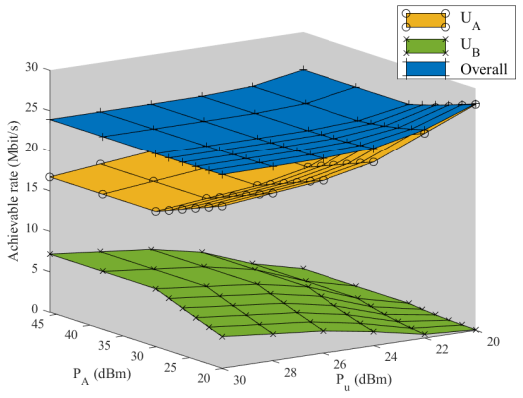
UT groups and thus generate interference increases with a larger radius, hence increasing the computational time. For this reason, $r_{\mathrm{A}} = r_{\mathrm{B}} = 90$ m will be used from here in the simulations.

B. MA-QL Performance

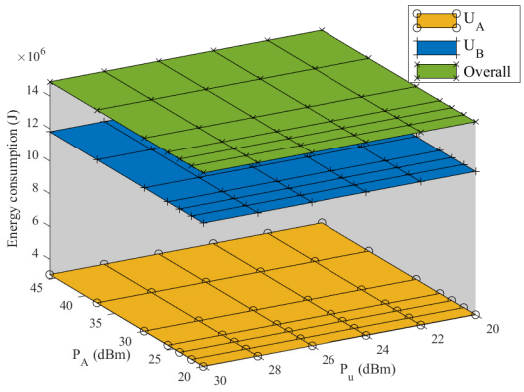
Fig. 6 shows the performance of the proposed algorithm, in terms of achievable rate, energy consumption, and energy efficiency with respect to the transmit power P_{A} and P_u. According to the results in Fig. 6(a), we can observe that the downlink achievable rate decreases while the uplink achievable rate increases with the rise of P_u. The contrary is noticed when P_{A} increases. This is because the uplink surface growth is due to the positive correlation between the uplink achievable rate and P_u, while the downlink surface growth is due to the positive correlation between the downlink achievable rate and P_{A}. The decrease of the surfaces is due to the increase of uplink-downlink interference. Specifically, $\mathrm{Int}^{\mathrm{dl}}$ increases as P_u increases and negatively affects the achievable rate in the downlink, while $\mathrm{Int}^{\mathrm{ul}}$ negatively affects the uplink achievable rate as P_{A} increases.

Fig. 6(b) demonstrates the results for the energy consumption. The analytical surfaces are obtained from (26), (29), and (30). Results illustrate that the transmit power does not particularly affect energy consumption, which is mainly due to flight and hovering consumption. An interesting phenomenon is observed that $\mathrm{U_B}$ consumes more energy than $\mathrm{U_A}$, even if it does not have communication consumption. The behavior is explainable by the uplink lower achievable rate, which extends the hovering time in the $\mathrm{U_B}$ service and makes it consume more to hover and collect data.

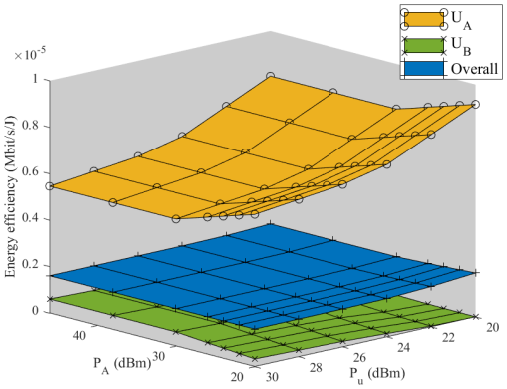
Furthermore, Fig. 6(c) provides the results for the energy efficiency in the UAV-enabled network. The analytical surfaces are obtained from (31) using the only relevant variables of the corresponding UAVs. The results illustrate a lower $\mathrm{U_B}$ energy efficiency, which also decreases the overall system efficiency. More importantly, it is shown that the optimal



(a) Achievable rate.



(b) Energy consumption.



(c) Energy efficiency.

Fig. 6: The performance of the proposed MA-QL algorithm.

energy efficiency point is with maximum P_{A} and minimum P_u.

Fig. 7 demonstrates the influence of $\mathrm{U_A}$ and $\mathrm{U_B}$ altitude on the overall system achievable rate. The solid curves have been obtained by the sum of (18) and (23). The results illustrate the constant decrease of the achievable rate as the altitude increases. More importantly, it shows a higher influence of the altitude variation of H_{A} on the overall system achievable rate, and a greater curve decrease can be noted. Finally, Fig. 8 gives the holistic optimized trajectories of $\mathrm{U_A}$ and $\mathrm{U_B}$, respectively.

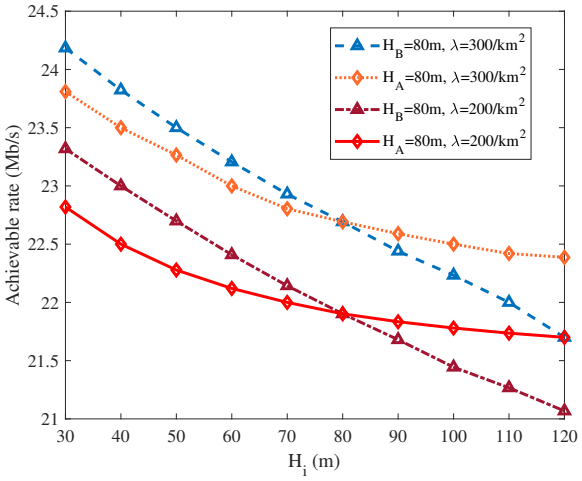


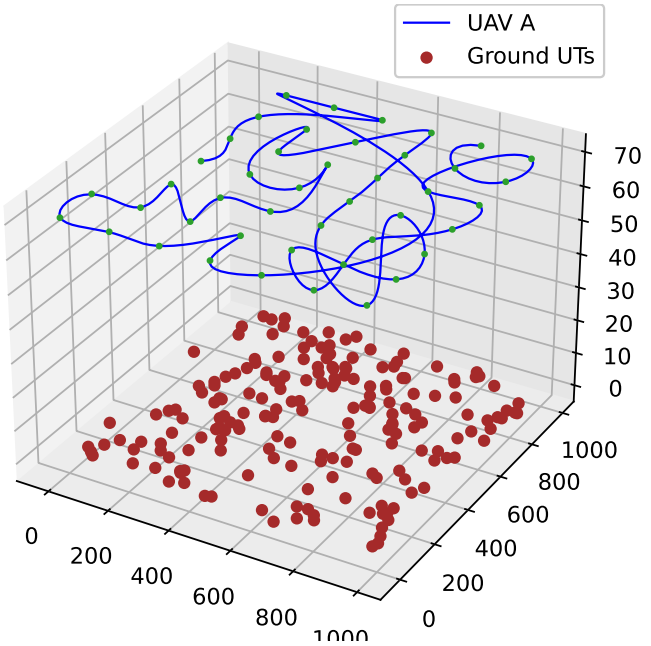
Fig. 7: The achievable rate against altitude variation.

C. Performance Comparison

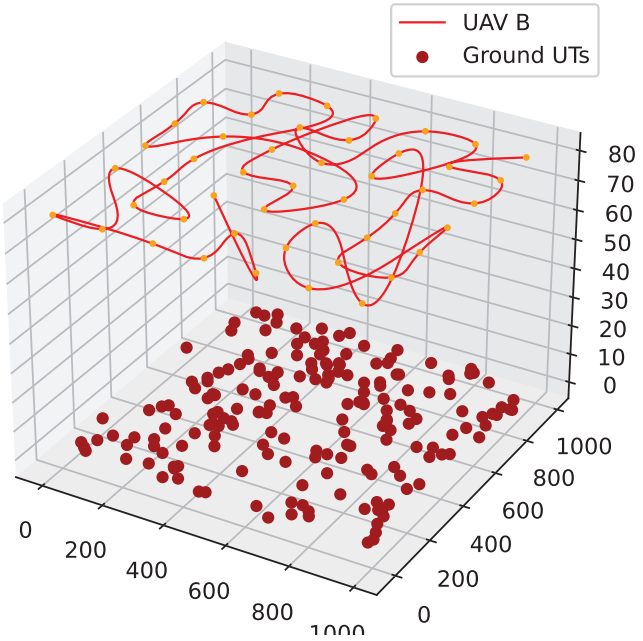
In this sub-section, we compare our algorithm with two other approaches, i.e., a random action selection [48] and a zigzag trajectory approach [49]. Fig. 9 provides the results for the overall system results in terms of achievable rate, energy consumption, and energy efficiency of the three approaches, against multiple densities. The achievable rate curves in Fig. 9(a) are obtained by the sum of (18) and (23), the energy consumption curves in 9(b), by the sum of (26), (29), and (30), while the energy efficiency curves in Fig. 9(c), by (31). Results in Fig. 9(a) illustrate that with any kind of UT density the MA-QL algorithm and the zigzag approach overcome the random approach. Results in Fig. 9(b) illustrate that the MA-QL algorithm and the random approach with any kind of UT density can save more than the zigzag approach. More importantly, Fig. 9(c) demonstrates the superiority of energy efficiency of the MA-QL algorithm compared to the cited approaches, with the optimal point at $\lambda = 50$ UTs/km^2. Finally, Fig. 10 compares the trajectory. In all three cases, the UAVs totally cover the UTs distributed on the ground. But it can be easily noticed that the more confusing trajectory in Fig. 10(c) does not allow to reach good levels of achievable rate and the zigzag trajectory in Fig. 10(b) is energy-intensive due to square direction changes.

VII. Conclusion

This paper considered an edge network where two UAVs collaboratively provide service for uplink and downlink transmission. The stochastic geometry theory and queue theory tools have been introduced for performance analysis and the exact and lower bound expressions for the average achievable rate for downlink and uplink were derived. Then, aiming to optimize the energy efficiency of the whole system, the MA-QL algorithm was proposed with the improved parameters including the learning rate and ϵ factor. Our analysis has shown that the proposed MA-QL algorithm achieved much better performance than the zigzag and random trajectory based approaches.



(a) Trajectory of $\mathrm{U_A}$.



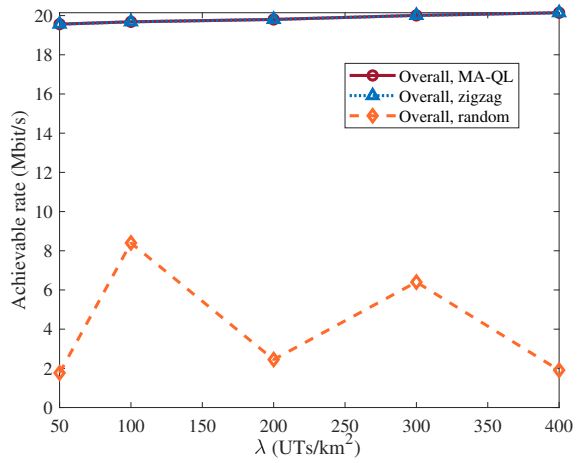
(b) Trajectory of $\mathrm{U_B}$.

Fig. 8: Optimized trajectory using the proposed algorithm.

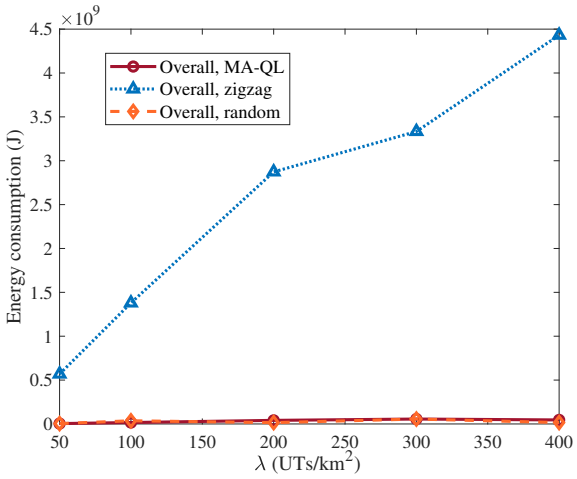
Appendix A: Proof of Theorem 1

The conditional average downlink achievable rate of the typical UT u served by $\mathrm{U_A}$ is expressed as

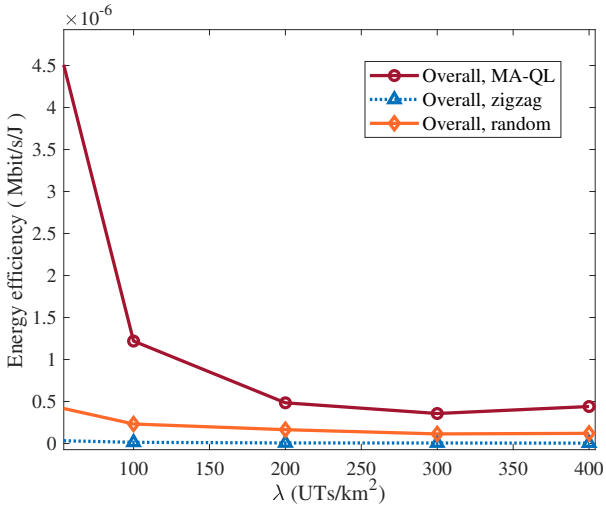
$$\bar{R}_{\mathrm{A}} = \frac{2\pi}{\ln 2} \int_0^{r_{\mathrm{A}}} \int_0^{\infty} \frac{F_{\gamma_{\mathrm{A}}}(\chi, x)}{1+\chi} x d\chi dx, \qquad \text{(A.1)}$$



(a) Achievable rate.

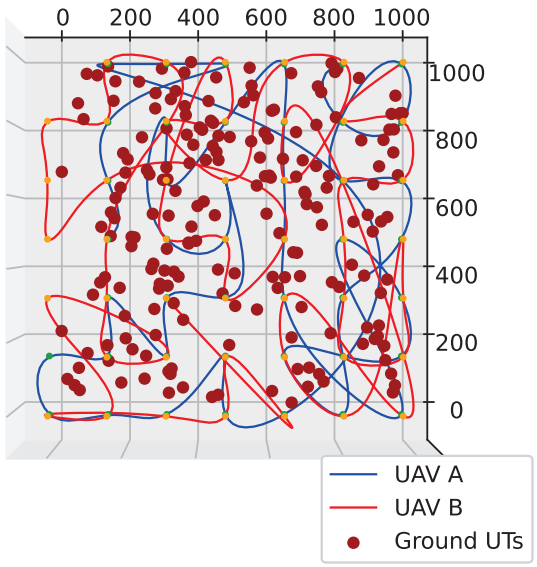


(b) Energy consumption.

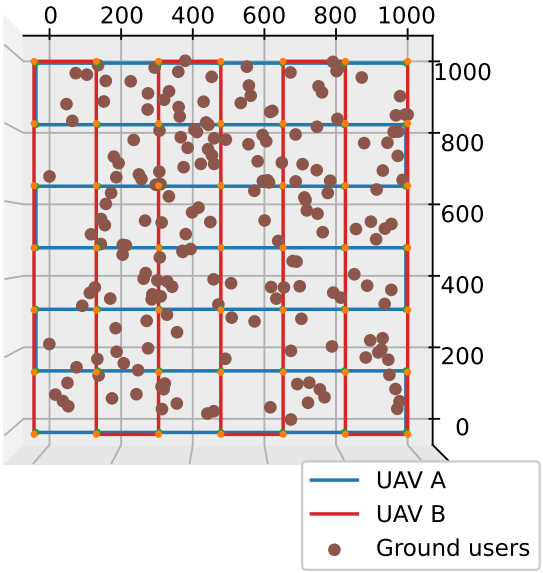


(c) Energy efficiency.

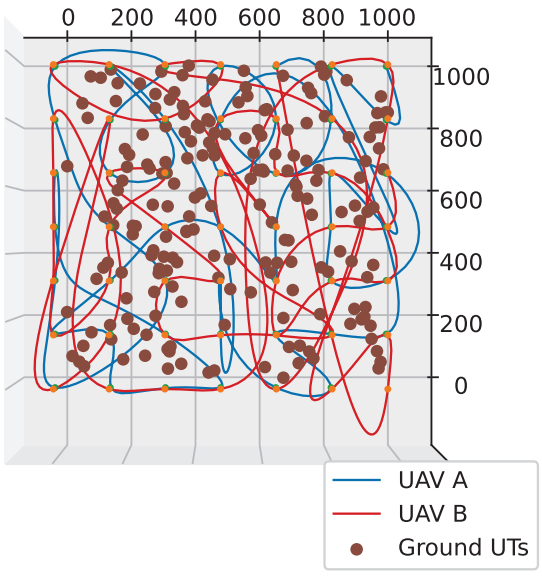
Fig. 9: Comparison of the proposed MA-QL algorithm, the random trajectory based, and the zigzag trajectory based approaches.



(a) Optimized trajectory using the proposed algorithm.



(b) Zigzag trajectory.



(c) Random trajectory.

Fig. 10: UAV trajectory under different schemes.

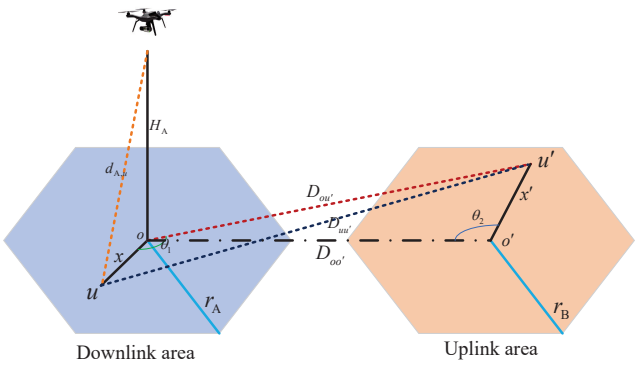


Fig. 11: Distance diagram between $\mathrm{U_A}$ and $\mathrm{U_B}$.

where F_{γ_A} is the complementary cumulative distribution function (CCDF) of γ_A in (3) and is given by, according to [50],

$$F_{\gamma_\mathrm{A}} = \underbrace{\Pr\left[\frac{P_\mathrm{A}\beta_0 h_{\mathrm{A},u} d_\mathrm{A}^{-\alpha^\mathrm{L}}}{\delta \mathrm{Int}^{\mathrm{dl}} + \sigma^2} > \chi\right]}_{\Pi_\mathrm{A}^\mathrm{L}(\chi,x)} \Pr^\mathrm{L}(x) + \underbrace{\Pr\left[\frac{P_A\beta_0 h_{\mathrm{A},u} d_\mathrm{A}^{-\alpha^\mathrm{N}}}{\delta \mathrm{Int}^{\mathrm{dl}} + \sigma^2} > \chi\right]}_{\Pi_\mathrm{A}^\mathrm{N}(\chi,x)} \Pr^\mathrm{N}(x), \quad \text{(A.2)}$$

where $\Pi_\mathrm{A}^\mathrm{L}(\chi,x)$ and $\Pi_\mathrm{A}^\mathrm{N}(\chi,x)$ are computed as (13) and (14), respectively, and Ψ_1 is obtained based on (5), i.e.,

$$\Psi_1 = \mathbb{E}\{P_u\beta_0 g_u \bar{d}_{u,\Phi_\mathrm{B}}^{-\alpha^\mathrm{N}}\} = P_u\beta_0 \bar{d}_{u,u'}^{-\alpha^\mathrm{N}}. \quad \text{(A.3)}$$

where $\bar{d}_{u,u'}$ is the average distance between the uplink typical user u' to the downlink typical user u, as shown in Fig. 11, and is given by

$$\bar{d}_{u,u'} = \frac{\int_0^{2\pi}\int_0^{r_\mathrm{A}} \Xi(d_{\mathrm{A},u'}, x, \theta_1) x \mathrm{d}x\mathrm{d}\theta_1}{\pi r_\mathrm{A}^2}, \quad \text{(A.4)}$$

according to the cosine theorem in (16). Similarly, the average distance $d_{\mathrm{A},u'}$ between the downlink typical user u to the centre o' of a certain subarea $k \in \mathcal{K}_\mathrm{B}$ is computed. Finally, the proof is completed by replacing $d_{\mathrm{A},u'}$ with $\Delta\bar{D}$.

Appendix B: Proof of Theorem 2

We first rewrite the projection distance x using the polar coordinates. According to the Jensen's inequality, the lower bound for $\bar{R}_\mathrm{A}$ is computed as

$$\bar{R}_\mathrm{A}^{\mathrm{lb}} = \frac{1}{\pi r_\mathrm{A}^2}\int_0^{2\pi}\int_0^{r_\mathrm{A}} \log_2\left(1 + \frac{1}{\mathbb{E}\{\gamma_\mathrm{A}^{-1}(x,\theta_1)\}}\right) x\mathrm{d}x\mathrm{d}\theta_1. \quad \text{(B.1)}$$

Then $\mathbb{E}\{\gamma_\mathrm{A}^{-1}(x,\theta_1)\}$ can be computed as [50]

$$\mathbb{E}\{\gamma_\mathrm{A}^{-1}(x,\theta_1)\} = \frac{\delta\Psi_1 + \sigma^2}{P_\mathrm{A} d_\mathrm{A}^{-\alpha_\mathrm{L}} \Pr^\mathrm{L} + P_\mathrm{A} d_\mathrm{A}^{-\alpha_\mathrm{N}} \Pr^\mathrm{N}}. \quad \text{(B.2)}$$

where Ψ_1 is given in (A.3). Thus, we obtain (18) in ***Theorem 2*** and complete the proof.

Appendix C: Proof of Theorem 3

Similar to the proof of ***Theorem 1***, the conditional average uplink achievable rate for a typical user u' served by $\mathrm{U_B}$ is expressed as

$$\bar{R}_\mathrm{B} = \frac{2\pi}{\ln 2}\int_0^{r_\mathrm{B}}\int_0^{\infty} \frac{F_{\gamma_\mathrm{B}}(\chi,x')}{1+\chi} x' \mathrm{d}\chi\mathrm{d}x', \quad \text{(C.1)}$$

where $F_{\gamma_\mathrm{B}}(\chi,x')$ is the CCDF of γ_B in (7) and is given by

$$F_{\gamma_\mathrm{B}}(\chi,x') = \underbrace{\Pr\left[\frac{P_u\beta_0 h_{\mathrm{B},u} d_\mathrm{B}^{-\alpha^\mathrm{L}}(x')}{\delta \mathrm{Int}^{\mathrm{ul}} + \sigma^2} > \chi\right]}_{\Pi_\mathrm{B}^\mathrm{L}(\chi,x')} \Pr^\mathrm{L}(x') + \underbrace{\Pr\left[\frac{P_u\beta_0 h_{\mathrm{B},u} d_\mathrm{B}^{-\alpha^\mathrm{N}}(x')}{\delta \mathrm{Int}^{\mathrm{ul}} + \sigma^2} > \chi\right]}_{\Pi_\mathrm{B}^\mathrm{N}(\chi,x')} \Pr^\mathrm{N}(x'), \quad \text{(C.2)}$$

where $\Pi_\mathrm{B}^\mathrm{L}$ and $\Pi_\mathrm{B}^\mathrm{N}$ are given in (21) and (22), respectively. Note that only the LoS link is between $\mathrm{U_A}$ and $\mathrm{U_B}$ and the average interference is obtained as $\mathbb{E}(\mathrm{Int}^{\mathrm{ul}}) = \Psi_2 = P_\mathrm{A}\beta_0 d_{\mathrm{A,B}}^{-\alpha^\mathrm{L}}$. Therefore, we obtain (20) in ***Theorem 3*** and complete the proof.

Appendix D: Proof of Theorem 4

Using the Jensen inequality again, a tractable lower bound for the conditional average uplink achievable rate $\bar{R}_\mathrm{B}$ is given as

$$\bar{R}_\mathrm{B}^{\mathrm{lb}} = \frac{1}{\pi r_\mathrm{B}^2}\int_0^{2\pi}\int_0^{r_\mathrm{B}} \log_2\left(1 + \frac{1}{\mathbb{E}\{\gamma_\mathrm{B}^{-1}(x',\theta_2)\}}\right) x'\mathrm{d}x'\mathrm{d}\theta_2. \quad \text{(D.1)}$$

where x' is the projection distance between an uplink typical user u' to the central of a ceratin sub-area and $\mathbb{E}\{\gamma_\mathrm{B}^{-1}(x',\theta_2)\}$ is calculated as

$$\mathbb{E}\{\gamma_\mathrm{B}^{-1}(x',\theta_2)\} = \frac{\delta P_\mathrm{A}\Psi_2 + \sigma^2}{P_u\beta_0\left(\Pr^\mathrm{L} d_\mathrm{B}^{-\alpha_\mathrm{L}} + \Pr^\mathrm{L} d_\mathrm{B}^{-\alpha^\mathrm{N}}\right)}. \quad \text{(D.2)}$$

Hence, we obtain (23) and (24) in ***Theorem 4*** and complete the proof.

References

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022 white paper," Cisco, Tech. Rep., Feb. 2019. [Online]. Available: https://s3.amazonaws.com/media.mediapost.com/uploads/CiscoForecast.pdf

[2] T. Alladi, Naren, G. Bansal, V. Chamola, and M. Guizani, "SecAuthUAV: A novel authentication scheme for UAV-ground station and UAV-UAV communication," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15 068–15 077, Dec. 2020.

[3] C. M. Gevaert, J. Suomalainen, J. Tang, and L. Kooistra, "Generation of spectral-temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3140–3146, June 2015.

[4] M. Bacco, A. Berton, A. Gotta, and L. Caviglione, "IEEE 802.15.4 air-ground UAV communications in smart farming scenarios," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1910–1913, Sept. 2018.

[5] K. Feng, W. Li, S. Ge, and F. Pan, "Packages delivery based on marker detection for UAVs," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Aug. 2020, pp. 2094–2099.

[6] M. Asadpour, B. Van den Bergh, D. Giustiniano, K. A. Hummel, S. Pollin, and B. Plattner, "Micro aerial vehicle networks: An experimental analysis of challenges and opportunities," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 141–149, July 2014.

[7] L. Gupta, R. Jain, and G. Vaszkun, "Survey of Important Issues in UAV Communication Networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1123–1152, Nov. 2016.

[8] C. H. Liu, Z. Chen, J. Tang, J. Xu, and C. Piao, "Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2059–2070, Sept. 2018.

[9] N. Gao, X. Li, S. Jin, and M. Matthaiou, "3-D deployment of UAV swarm for massive MIMO communications," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3022–3034, 2021.

[10] C. K. Armeniakos, P. S. Bithas, and A. G. Kanatas, "SIR analysis in 3D UAV networks: A stochastic geometry approach," *IEEE Access*, vol. 8, pp. 204 963–204 973, Nov. 2020.

[11] K. Yoshikawa, K. Yamamoto, T. Nishio, and M. Morikura, "Grid-based exclusive region design for 3D UAV networks: A stochastic geometry approach," *IEEE Access*, vol. 7, pp. 103 806–103 814, July 2019.

[12] C.-H. Liu, K.-H. Ho, and J.-Y. Wu, "Mmwave UAV networks with multi-cell association: Performance limit and optimization," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2814–2831, Dec. 2019.

[13] S. Zhang, J. Liu, and W. Sun, "Stochastic geometric analysis of multiple unmanned aerial vehicle-assisted communications over Internet of things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5446–5460, June 2019.

[14] V. V. Chetlur and H. S. Dhillon, "Downlink coverage analysis for a finite 3-D wireless network of unmanned aerial vehicles," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4543–4558, Oct. 2017.

[15] T. Hou, Y. Liu, Z. Song, X. Sun, and Y. Chen, "Multiple antenna aided NOMA in UAV networks: A stochastic geometry approach," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1031–1044, Feb. 2019.

[16] H. He, S. Zhang, Y. Zeng, and R. Zhang, "Joint altitude and beamwidth optimization for UAV-enabled multiuser communications," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 344–347, Feb. 2018.

[17] Y. Wu, W. Yang, X. Guan, and Q. Wu, "Energy-efficient trajectory design for UAV-enabled communication under malicious jamming," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 206–210, Feb. 2021.

[18] Z. Yang, W. Xu, and M. Shikh-Bahaei, "Energy efficient UAV communication with energy harvesting," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1913–1927, Feb. 2020.

[19] S. Fu, Y. Tang, Y. Wu, N. Zhang, H. Gu, C. Chen, and M. Liu, "Energy-efficient UAV enabled data collection via wireless charging: A reinforcement learning approach," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 10 209–10 219, June 2021.

[20] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3747–3760, June 2017.

[21] J. Chakareski, S. Naqvi, N. Mastronarde, J. Xu, F. Afghah, and A. Razi, "An energy efficient framework for UAV-assisted millimeter wave 5g heterogeneous cellular networks," *IEEE Trans. Green Comm. Netw.*, vol. 3, no. 1, pp. 37–44, Mar. 2019.

[22] C. Liu, W. Feng, J. Wang, Y. Chen, and N. Ge, "Aerial small cells using coordinated multiple UAVs: An energy efficiency optimization perspective," *IEEE Access*, vol. 7, pp. 122 838–122 848, Aug. 2019.

[23] N. I. Mowla, N. H. Tran, I. Doh, and K. Chae, "AFRL: Adaptive federated reinforcement learning for intelligent jamming defense in FANET," *Journal of Communications and Networks*, vol. 22, no. 3, pp. 244–258, June 2020.

[24] S. Zhu, L. Gui, N. Cheng, F. Sun, and Q. Zhang, "Joint design of access point selection and path planning for UAV-assisted cellular networks," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 220–233, Jan. 2020.

[25] Q. Liu, L. Shi, L. Sun, J. Li, M. Ding, and F. Shu, "Path planning for UAV-mounted mobile edge computing with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5723–5728, May 2020.

[26] L. Li, Q. Cheng, K. Xue, C. Yang, and Z. Han, "Downlink transmit power control in ultra-dense UAV network based on mean field game and deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15 594–15 605, Dec. 2020.

[27] U. Challita, W. Saad, and C. Bettstetter, "Interference Management for Cellular-Connected UAVs: A Deep Reinforcement Learning Approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2125–2140, Apr. 2019.

[28] J. Hu, H. Zhang, L. Song, R. Schober, and H. V. Poor, "Cooperative internet of UAVs: Distributed trajectory design by multi-agent deep reinforcement learning," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6807–6821, Nov. 2020.

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

[30] J. Xu, Q. Guo, L. Xiao, Z. Li, and G. Zhang, "Autonomous decision-making method for combat mission of UAV based on deep reinforcement learning," in *IEEE 4th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, vol. 1, Dec. 2019, pp. 538–544.

[31] H. Qi, Z. Hu, H. Huang, X. Wen, and Z. Lu, "Energy efficient 3-D UAV control for persistent communication service and fairness: A deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 53 172–53 184, Mar. 2020.

[32] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (UAVs) for energy-efficient Internet of things communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7574–7589, Nov. 2017.

[33] D. Hu, Q. Zhang, Q. Li, and J. Qin, "Joint position, decoding order, and power allocation optimization in UAV-based NOMA downlink communications," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2949–2960, June 2020.

[34] F. Jiang and A. L. Swindlehurst, "Optimization of UAV heading for the ground-to-air uplink," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 5, pp. 993–1005, June 2012.

[35] W. Mei, Q. Wu, and R. Zhang, "Cellular-connected UAV: Uplink association, power control and interference coordination," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5380–5393, Aug. 2019.

[36] L. Zhou, Z. Yang, S. Zhou, and W. Zhang, "Coverage probability analysis of uav cellular networks in urban environments," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2018, pp. 1–6.

[37] Y. Zhu, L. Wang, K.-K. Wong, S. Jin, and Z. Zheng, "Wireless power transfer in massive MIMO-aided HetNets with user association," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4181–4195, Oct. 2016.

[38] U. Bhat, *An Introduction to Queueing Theory: Modeling and Analysis in Applications*. Birkhäuser Boston, 2015.

[39] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329–2345, Apr. 2019.

[40] N. Gao, Y. Zeng, J. Wang, D. Wu, C. Zhang, Q. Song, J. Qian, and S. Jin, "Energy model for UAV communications: Experimental validation and model generalization," *China Commun.*, vol. 18, no. 7, pp. 253–264, 2021.

[41] J. J.Craig, *Introduction to Robotics, Mechanics and Control*. Upper Saddle River, NJ 07458: Pearson Education International, 2005.

[42] C. Zhan and Y. Zeng, "Energy-efficient data uploading for cellular-connected UAV systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7279–7292, Nov. 2020.

[43] A. Nowé, P. Vrancx, and Y.-M. De Hauwere, *Game Theory and Multi-agent Reinforcement Learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 441–470.

[44] C. H. Liu, X. Ma, X. Gao, and J. Tang, "Distributed energy-efficient multi-UAV navigation for long-term communication coverage by deep reinforcement learning," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1274–1285, June 2020.

[45] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, May 2019.

[46] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[47] Y. Zhu, G. Zheng, and M. Fitch, "Secrecy rate analysis of UAV-enabled mmwave networks using matérn hardcore point processes," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1397–1409, July 2018.

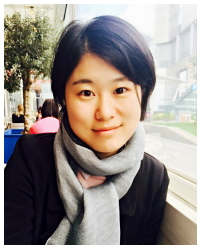
[48] X. Liu, M. Chen, and C. Yin, "Optimized trajectory design in UAV based cellular networks for 3D users: A double Q-learning approach," *IEEE J. Comm. and Inf. Netw.*, vol. 4, no. 1, pp. 24–32, Mar. 2019.

[49] J. Yu, Y. Zhu, H. Zhao, R. Cepeda-Lopez, T. Dagiuklas, and Y. Gao, "Dynamic coverage path planning of energy optimization in UAV-enabled edge computing networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Nanjing, China, Mar. 2021, pp. 1–6.

[50] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, 2015.

Wenchao Xia (Member, IEEE) received his B.S. degree in communication engineering and Ph.D. degree in communication and information systems from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2014 and 2019, respectively. From 2019 to 2020, he was a Postdoctoral Research Fellow with Singapore University of Technology and Design, Singapore. He is currently with the faculty of the Jiangsu Key Laboratory of Wireless Communications, College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interests include edge intelligence and multi-antenna communications.

He was a recipient of the IEEE Globecom Best Paper Award in 2016 and the IEEE JC&S Best Paper Award in 2021. He serves as an Associate Editor for the IET Electronics Letters.



Yongxu Zhu (Senior Member, IEEE) received the Ph.D degree in Electrical Engineering from University College London in 2017. From 2017 to 2019, she was a Research Associate with Loughborough University. She is currently a Senior Lecturer with the Division of Computer Science and informatics, London South Bank University. She has served as an Editor of the IEEE Wireless Communication Letters. Her research interests include future wireless communication, heterogeneous networks, and physical-layer security.

Lorenzo De Simone received the B.S. degree in Management Engineering from La Sapienza University of Rome, Italy, in 2019, and the M.Sc. degree in Data Science from the London South Bank University, U.K., in 2020, where he is currently pursuing the Ph.D. degree. His research interests include Machine learning, Mathematical Optimization, and UAV applications for Wireless Communications.

Prof. Tasos Dagiuklas (Member, IEEE) is a leading researcher andexpert in the fields of smart Internet technologies. He is the leader of the Smart Internet Technologies(SuITE) research group at the London South Bank University where he also acts as the Head of Cognitive Systems Research Centre. Tasos received the Engineering Degree from the University of Patras-Greece in 1989, the M.Sc. from the University of Manchester- UK in 1991 and the Ph.D. from theUniversity of Essex-UK in 1995, all in Electrical Engineering. He has been a principal investigator,co-investigator, project and technical manager, coordinator and focal person of more than 25 internationally R&D and Capacity training projects in theareas of Fixed- Mobile Convergence, 4G/5G networking technologies, VoIP and multimedia networking. His research interests lie in the field of SystemsBeyond 5G/6G networking technologies, programmable networks, UAVs, V2X communications and cyber security for smart Internet systems.

Kai-Kit Wong (Fellow, IEEE) received the B.Eng., M.Phil., and Ph.D. degrees in electrical and electronic engineering from The Hong Kong University of Science and Technology, Hong Kong, in 1996, 1998, and 2001, respectively. After graduation, he took up academic and research positions at The University of Hong Kong, Lucent Technologies, Bell-Labs, Holmdel, the Smart Antennas Research Group, Stanford University, and the University of Hull, U.K. He is the Chair of wireless communications with the Department of Electronic and Electrical Engineering, University College London, U.K. His current research interests include around 5G and beyond mobile communications, including topics such as massive multiple-input multiple-output, full-duplex communications, millimeter-wave communications, edge caching and fog networking, physical layer security, wireless power transfer and mobile computing, V2X communications, and, of course, cognitive radios. There are also a few other unconventional research topics that he has set his heart on, including, for example, fluid antenna communications systems, remote ECG detection, and so on. He was a co-recipient of the 2013 IEEE Signal Processing Letters Best Paper Award and the 2000 IEEE VTS Japan Chapter Award from the IEEE Vehicular Technology Conference, Japan, in 2000, and a few other international best paper awards. He is a Fellow of IET and is also on the editorial board of several international journals. He has been serving as a Senior Editor for the IEEE COMMUNICATIONS LETTERS, since 2012, and also for the IEEE WIRELESS COMMUNICATIONS LETTERS, since 2016. He has been an Area Editor for Wireless Communication Theory and Systems I of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, since 2018. He served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2005 to 2011, and as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS, from 2009 to 2012. He was also a Guest Editor of the IEEE JSAC Special Issue on Virtual MIMO, in 2013. He is a Guest Editor of the IEEE JSAC Special Issue on Physical Layer Security for 5G.

Gan Zheng (Fellow, IEEE) received the BEng and the MEng from Tianjin University, Tianjin, China, in 2002 and 2004, respectively, both in Electronic and Information Engineering, and the PhD degree in Electrical and Electronic Engineering from The University of Hong Kong in 2008. He is currently Professor of Signal Processing and Wireless Communications in the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, UK. His research interests include machine learning for communications, UAV communications, mobile edge caching, full-duplex radio, and wireless power transfer. He is the first recipient for the 2013 IEEE Signal Processing Letters Best Paper Award, and he also received 2015 GLOBECOM Best Paper Award, and 2018 IEEE Technical Committee on Green Communications & Computing Best Paper Award. He was listed as a Highly Cited Researcher by Thomson Reuters/Clarivate Analytics in 2019 and he is an IEEE Fellow. He currently serves as an Associate Editor for IEEE Communications Letters and IEEE Wireless Communications Letters.