The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 10 (2024) 15–21

www.elsevier.com/locate/icte

On sum-rate maximization in downlink UAV-aided RSMA systems

Duc-Thien Hua[a], Quang Tuan Do[a], Nhu-Ngoc Dao[b,*], Sungrae Cho[a,*]

[a] School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea
[b] Faculty of Computer Science, Ho Chi Minh City Open University, Ho Chi Minh City 70000, Vietnam

Received 6 December 2022; received in revised form 3 February 2023; accepted 8 March 2023
Available online 16 March 2023

Abstract

The synergy of uncrewed aerial vehicles (UAVs) and the rate-splitting multiple access (RSMA) technique has thrived as a crucial enabler for multi-user broadband sixth-generation networks. This paper investigates a UAV-aided RSMA downlink communication system considering user mobility and the uniform rectangular array antenna design. In particular, the sum-rate maximization problem is formulated where the UAV beamforming matrix, common rate allocation, and UAV trajectory design are jointly optimized. The deep deterministic policy gradient (DDPG) approach has been proposed to address the nonconcave objective function. In addition, the safe action shaping technique is incorporated into the algorithm to satisfy the variable constraints of the problem statement. The numerical simulation results demonstrated that the proposed approach outperformed other benchmark schemes in various settings of transmit powers and fading environment levels.

Keywords: Rate-splitting multiple access (RSMA); Uncrewed aerial vehicle (UAV) trajectory; Deep reinforcement learning; User mobility

1. Introduction

To cope with the unprecedented growth of the Internet of Things (IoT) paradigms and their applications, the integration of uncrewed aerial vehicles (UAVs) into mobile ecosystems has increasingly thrived owing to their dynamic and quick implementation [1]. In addition, another leading advantage of UAV systems is that such platforms are becoming progressively smaller and consuming less energy, significantly lowering the cost of manufacturing and installation [2]. Thus, the UAV has been considered as one of the most propitious enablers to complement the three-dimensional (3D) wireless communication architecture envisioned for the sixth-generation (6G) networks [3]. In this context, UAV acts as an aerial base station (BS) to provide line of sight (LoS) wireless channels to ground IoT devices, leading to *UAV-aided communication systems* [4]. Numerous studies have investigated the benefits of these systems. For instance, Hua et al. designed the optimized trajectory of multiple UAVs to provide efficient downlink communications [5]. In [6], Yeom et al. investigated a multi-UAV-aided downlink NOMA network with a virtual full-duplex proposition to address the outage probability metrics. Nonetheless, none of the aforementioned works addressed the state-of-the-art rate-splitting multiple access (RSMA) technique, which was proven to outperform NOMA regarding various metrics [7–9].

In particular, the RSMA mechanism is an efficient access technique that effectively exploits the advantages of multiple antenna designs for massive IoT device scenarios [10]. In particular, according to the numerical results in [8], RSMA can handle the curse of mobility expected in the sixth-generation (6G) networks. Ahmad et al. studied the coverage of the cloud-radio access RSMA network, in which the UAV is an aerial BS providing downlink communication [11]. However, the UAV was set in a fixed location, instead of being studied while efficiently traveling and hovering. Jaafar et al. studied the UAV-aided RSMA downlink communication, in which the joint variables of the precoding matrix, common rate vector, and UAV location are optimized to maximize the weighted sum rate of the system [12]. Bastami et al. investigated the secrecy max–min fairness rate considering an imperfect channel and eavesdropper in UAV-aided cooperative RSMA downlink communication [13].

Nevertheless, these previous studies considered the dominant LoS channel model, which is believed to be an impractical and obsolete model. In addition, the user mobility model has not been provided, particularly in these existing

* Corresponding authors.
E-mail addresses: thien@uclab.re.kr (D.-T. Hua), dqtuan@uclab.re.kr (Q.T. Do), ngoc.dn@ou.edu.vn (N.-N. Dao), srcho@cau.ac.kr (S. Cho).

Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2023.03.001

studies. Furthermore, the mentioned studies applied conventional optimization methods, such as alternative and convex optimizations. These approaches are immensely complex and computationally burdensome, and they are workable in low-dimensional environment settings with prior knowledge of the channel state information (CSI). In contrast, deep reinforcement learning (DRL)-based approaches are capable of extracting the features of the dynamic channel pattern, and determining the joint optimal action sets with a lower complexity and faster execution time [14]. Motivated by these observations, the main contributions of this study are summarized as follows:

- First, we investigate a UAV-aided RSMA downlink communication system, in which, the mobility functions and uniform rectangular array (URA) beamforming design of the UAV are considered.
- Second, we formulate the sum-rate maximization problem statement, where the UAV beamforming matrix, common rate allocation, and UAV trajectory design are jointly optimized. In addition, the formulated problem is transformed into a Markov decision process (MDP) framework as a DRL task. In this framework, the UAV interacts, observes, and learns the channel patterns without any prior CSI acknowledgment.
- Third, a DRL-inspired approach was proposed to address the optimization problem. In addition, a safe action-shaping technique was integrated into the proposed algorithm to satisfy all constraints.
- Fourth, we evaluated the performance of the proposed method and compared it with benchmark schemes to demonstrate the superiority of the proposed approach, especially with regard to the maximum sum-rate metrics.

The rest of the paper is organized as follows. Section 2 investigates the system model. Section 3 provides the problem formulation and transformation. Next, Section 5 presents the evaluation of the simulation performance of the proposed method in Section 4. Finally, Section 6 concludes the paper.

2. System model

In the studied scenario, a mobile network where a UAV, equipped with URA $N_1 \times N_2$ antennas, is employed to provide service to K ground IoT devices. As illustrated in Fig. 1, the UAV travels to facilitate the air-to-ground channel and transmit signals to ground users. In this scenario, the UAV and K users are progressively moving to mimic the dynamic practical scenario. For simple indication, $k = 0$ and $k \in [1, K]$ indicate the UAV and the k-th ground user, respectively. we propose the movement functions determined using the three independent variables, including the moving time δ, the UAV velocity $V_k(t) \in [0, V_k^{max}]$, and the directional angle $\varphi_k(t) \in [0, 2\pi]$ on the xy-plane representing the moving direction of the device. Each timeslot t comprises two intervals, including the moving and hovering intervals, where the hovering time is calculated by $|t - \delta|$. In particular, the UAV hovers to provide communication to the users after having traveled for δ s. The next location can be calculated as follows:

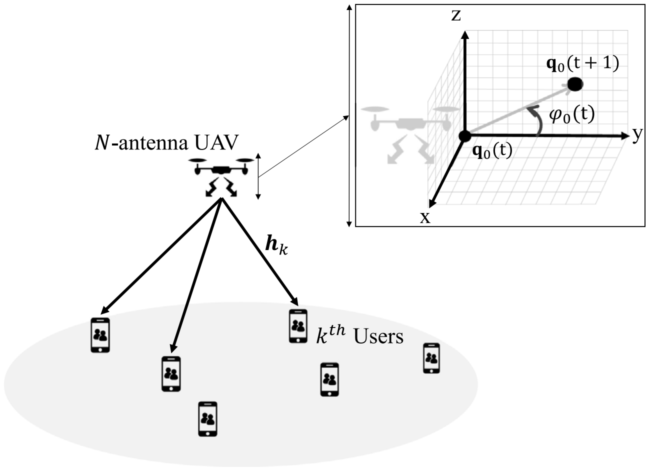


Fig. 1. Uncrewed aerial vehicle-aided rate-splitting multiple access downlink transmission system, where an aerial base station traverses to serve K ground users.

$$\begin{aligned} x_k(t+1) &= x_k(t) + \delta V_k(t)\cos(\varphi_k(t)), \\ y_k(t+1) &= y_k(t) + \delta V_k(t)\sin(\varphi_k(t)), \end{aligned} \tag{1}$$

where the velocity and the directional angle of the k-th user are uniformly generated from the distributions $\mathcal{CN}(V_k^{max}, 1)$ and $\mathcal{CN}(2\pi, 1)$, respectively.

In comparison with terrestrial communication settings, the air-to-ground links are dominated by the elevation angle and altitude of the aerial BS [15]. As in [16], the block-fading Rician model for the channel from the aerial BS to the k-th user is defined as

$$\mathbf{h}_k = L_{h_k}\left(\sqrt{\frac{\kappa}{\kappa+1}}\mathbf{h}_k^{LoS} + \sqrt{\frac{1}{\kappa+1}}\mathbf{h}_k^{NLoS}\right), \tag{2}$$

where κ is the Rician factor, $\mathbf{h}_k^{NLoS} \sim \mathcal{CN}(0, 1)$ is the NLoS component, and $\mathbf{h}_k^{LoS}$ is the LoS component, which is expressed as

$$\mathbf{h}_k^{LoS} = \boldsymbol{a}_k(\phi_k^e, \phi_k^a) = \text{vec}\left(\boldsymbol{a}_{N_1}(\rho)\boldsymbol{a}_{N_2}^T(\zeta)\right), \tag{3}$$

where $a_k(\cdot) \in \mathbb{C}^{N\times 1}$ is the steering array defined by the azimuth angle of departure ϕ_k^a and elevation angle of departure ϕ_k^e. We adopt the steering array as in [17], in which the operator vec$(\cdot)$ maps an $N_1 \times N_2$ matrix into an N-dimensional array by stacking the column elements of the matrix, i.e., $N_1 \times N_2 = N$. The $\mathbf{a}_{N_1}(\rho)$ and $\mathbf{a}_{N_2}(\zeta)$ are calculated as follows:

$$\begin{aligned} \boldsymbol{a}_{N_1}(\rho) &= [1, e^{j\rho}, \ldots, e^{j(N_1-1)\rho}]^T, \boldsymbol{a}_{N_2}(\zeta) \\ &= [1, e^{j\zeta}, \ldots, e^{j(N_2-1)\zeta}]^T, \end{aligned}$$

where

$$\begin{aligned} \rho &= \frac{2\pi}{\lambda} d_{N_1}\cos\phi_k^a \sin\phi_k^e = \frac{2\pi}{\lambda} d_{N_1}\frac{x_0 - x_k}{d_{0,k}}, \\ \zeta &= \frac{2\pi}{\lambda} d_{N_2}\sin\phi_k^a \sin\phi_k^e = \frac{2\pi}{\lambda} d_{N_2}\frac{y_0 - y_k}{d_{0,k}}, \end{aligned}$$

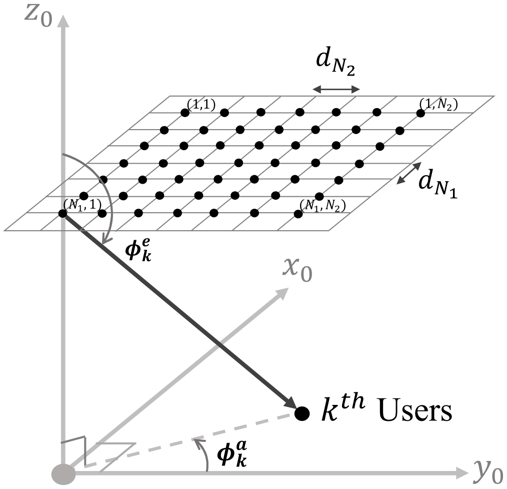


Fig. 2. Depiction of the steering array vector where the incoming signals are incidental from the uncrewed aerial vehicle to the k-th user.

in which λ is the wavelength, d_{N_1} and d_{N_2} are the respective distances between the two consecutive vertical, and horizontal antennas; $d_{0,k} = \sqrt{(x_0 - x_k)^2 + (y_0 - y_k)^2 + H^2}$ is the distance between the UAV and k-th user. In addition, Fig. 2 depicts the steering array of the incoming signal traverse from the UAV to the k-th user.

For the RSMA physical layer design, a robust downlink one-layer rate-splitting setting is studied. User-distinct messages are divided into common and private submessages. Then, all common messages of the K user are encoded into a single stream s_0. In addition, K private messages are encoded into separate private streams $\{s_k\}$. Thus, the stream vector to be transmitted is signified as $\mathbf{s}_k = [s^c, s_1^p, \ldots, s_K^p]^T$. Given $\mathbf{P} = [\mathbf{p}^c, \mathbf{p}_1^p \ldots, \mathbf{p}_K^p] \in \mathbb{C}^{N\times(K+1)}$ as the precoding matrix, the received signal at the k-th user is written as

$$y_k = \underbrace{\mathbf{h}_k^H \mathbf{p}^c s^c + \mathbf{h}_k^H \mathbf{p}_k^p s_k^p}_{\text{desired signal}} + \underbrace{\sum_{j\in\mathcal{K}, j\neq k} \mathbf{h}_k^H \mathbf{p}_j^p s_j^p}_{\text{interference signal}} + \sigma_k, \quad (4)$$

where σ_k is the additive white Gaussian noise. Furthermore, the constraint of the transmit power budget at the UAV follows $\sum_{k\in K} \|\mathbf{p}_k\|^2 \leq P_t$ where P_t is the maximum transmit power.

At the receiver, each user reconstructs its intended message by first decoding the common stream s^c by treating all private streams as noise. Then, the k-th user decodes its corresponding private stream s_k^p by treating the interference from the other private streams as noise. The decoded common and private messages are eventually combined. The corresponding signal-to-interference-plus-noise ratio of the common and private streams at the k-th user can be respectively obtained as follows:

$$\gamma_k^c = \frac{\left|\mathbf{h}_k^H \mathbf{p}^c\right|^2}{\sum_{j\in\mathcal{K}} \left|\mathbf{h}_j^H \mathbf{p}_j^p\right|^2 + {\sigma_j}^2}, \quad (5)$$

$$\gamma_k^p = \frac{\left|\mathbf{h}_k^H \mathbf{p}_k^p\right|^2}{\sum_{j\in\mathcal{K}, j\neq k} \left|\mathbf{h}_j^H \mathbf{p}_j^p\right|^2 + {\sigma_j}^2}. \quad (6)$$

The achievable rate of user k is calculated as $R_k^c = \log_2(1+\gamma_k^c)$ and $R_k^p = \log_2(1 + \gamma_k^p)$, respectively.

$$R_k^c = \log_2(1 + \gamma_k^c), \quad R_k^p = \log_2(1 + \gamma_k^p). \quad (7)$$

To ensure s^c is successfully decoded by all users, the actual rate of the common message R_c cannot exceed $R_c = min(R_1^c, \ldots, R_K^c)$. As s^c contains all common submessages of K users, the rate distribution of R_c should adapt to the number of sub-messages that each user contributes. By denoting C_k as the portion of R_c allocated to the k-th user for W_k^c, i.e., $\sum_{k\in\mathcal{K}} C_k = R_c$, the overall achievable rate of user k comprises the relevant part of the rate of s^c and the rate of s_k^p, which is mathematically expressed as $R_k = C_k + R_k^p$.

3. Problem formulation

In this study, the achievable sum-rate (ASR) maximization problem is formulated. The precoding matrix $\mathbf{P}$, common rate allocation $\mathbf{c} = [C_1, C_2, \ldots, C_K]$, and variables $\varphi_0(t)$ and V_0 for the UAV movement are jointly optimized with the objective of maximizing the ASR. The optimization can be mathematically formulated as follows:

$$\mathcal{P}1: \max_{\mathbf{P}(t), \mathbf{c}(t), \varphi_0(t), V_0(t)} \sum_{k=1}^{K} R_k(t), \quad (8a)$$

$$s.t. \sum_{k\in\mathcal{K}} \|\mathbf{p}_k\|^2 \leq P_t, \quad (8b)$$

$$\sum_{k\in\mathcal{K}} C_k \leq min(R_1^c, \ldots, R_K^c), \quad (8c)$$

$$R_k \geq R_{th}, \forall k \in \mathcal{K}, \quad (8d)$$

$$C_k \geq 0, \forall k \in \mathcal{K}, \quad (8e)$$

$$0 \leq \varphi_0 \leq 2\pi, \quad (8f)$$

$$0 \leq V_0 \leq V_{max}, \quad (8g)$$

where R_{th} is the minimum desired rate. The transmit power constraint is expressed as (8b). Constraint (8c) guarantees that the common stream is successfully decoded by all users. Constraints (8d) and (8e) ensure that the rate is a positive value and must exceed a threshold to satisfy the quality of service. Constraints (8f) and (8g) denote the value range of the directional angle and UAV velocity.

To solve the proposed problem, we transform (8) into an MDP framework including the state, action, reward, and policy. Then, the DRL-based algorithm can determine the optimal solutions for the proposed problem statement. Each compartment of the MDP framework is further described below.

(1) State: The locations of K users are insufficient enough for the agent to extract the features of the studied environment. Thus, we introduce a $\bar{x}_k$ as the calculation of the mean values of the locations of K users at each step. The $(2 + K * 2 + 1)$-dimensional state includes the x, y coordinates of the aerial BS, x, y coordinates of the K users, and the common rate of the last step.

(2) Action: After observing the state, the agent determines the joint actions of the precoding matrix $\mathbf{P}(t)$, the common

rate vector $\mathbf{c}(t)$, the directional angle $\varphi_0(t)$, and the UAV speed $V_0(t)$. The action space $\mathcal{A}$ is the combination of all the possible continuous values of these variables. The precoding element p_k is a complex value that comprises the real part parts (i.e., $p_k = \Re(p_{k,n}) + j\Im(p_{k,n})$).

(3) Reward: The agent determines actions to maximize the accumulative expected reward; thus, the reward function ($r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$) should be decided related to the objective function (8a). Thus, the objective is to maximize the ASR (i.e., $r(t) = \sum_{k=1}^{K} R_k(t)$).

4. Proposed approach

This section proposes a DRL-inspired approach with action shaping to satisfy the action constraints in (8a). First, the preliminaries of the deep deterministic policy gradient (DDPG) algorithm is introduced. Because the studied environmental problems comprise huge state space and action space dimensions, deep neural network (DNN) models are used as universal approximators to extract and learn features of the complex high-dimensional state space $\mathcal{S}$. Specifically, the DDPG approach parameterizes the policy $\mu(s|\theta^{\mu})$ and the critic network $Q(s, a|\theta^{Q})$ with the weight parameter sets θ^{μ} and θ^{Q}, respectively. To achieve the optimal policy $\mu^{*}(s|\theta^{\mu})$ that determines the optimal joint actions a^{*} to achieve the expected accumulative Q-value $Q^{*}(s, a|\theta^{Q})$, the Bellman equation can be solved, which is mathematically expressed as

$$Q(s(t), a(t)|\theta^{Q}) = \max_{a \in \mathcal{A}} \left[r(t) + \gamma Q(s(t+1), a(t+1)|\theta^{Q}) \right]. \tag{9}$$

A replay buffer D comprising the tuple of the state $s(t)$, action $a(t)$, reward $r(t)$ and next state $s(t+1)$ at each time slot is used for the sampled updating process. The tuple of experiences is uniformly sampled and stored, then input into the DNN models for training. In addition, the target actor network $\mu'(s|\theta^{\mu'})$ and target critic network $Q'\left(s(t), a(t)|\theta^{Q'}\right)$ are also employed to diminish the instability learning issue [18]. The overall mean squared bellman equation Q-value loss $L(\theta^{Q})$ is defined as

$$L(\theta^{Q}) = \mathbb{E}_{s_s(t), a_s(t), r_s(t) \sim B} \left[(Q(s_s(t), a_s(t)|\theta^{Q}) - y(t))^2 \right], \tag{10}$$

where $Q(s_s(t), a_s(t)|\theta^{Q})$ is the value of the chosen action $a_s(t)$ at state $s_s(t)$. In addition, $y(t)$ is defined as

$$y(t) = r_s(t) + \gamma Q'\left(s_s(t+1), \mu'(s_s(t+1)|\theta^{\mu'})|\theta^{Q'}\right). \tag{11}$$

Moreover, to enhance the exploration of the training sample, the continuous policy $\mu(s(t)|\theta^{Q})$ is modified as follows:

$$a(t) = \mu(s(t)|\theta^{Q}) + \mathcal{N}(t), \tag{12}$$

where $\mathcal{N}(t)$ is added noise to ensure the exploration of the current policy. Such noise is generated based on the Ornstein Uhlenbeck process as in [19]. Nevertheless, the additional Ornstein Uhlenbeck noise induces the violation of Constraint (8). We use the Sigmoid activation to scale the actor output in the range of [0, 1] (i.e., $0 \leq \mu(s|\theta^{Q}) \leq 1$) and introduce an action shaping function that rescales the true value of $\mathbf{P}(t)$, $\mathbf{c}(t)$, $\varphi_0(t)$ and $V_0(t)$. For notational simplicity, $\bar{a}$ denotes the action directly output from $\boldsymbol{\mu}(s|\theta^{\mu})$ followed by the Sigmoid activation, whereas a is the action being scaled to the true value.

In terms of the beamforming action, redefine the k-th beamforming vector, $\mathbf{p}_k$ is redefined as $\bar{\mathbf{p}}_k = [\alpha_1, \alpha_2, \dots \alpha_{2N-1}, \alpha_N]^T \in \mathbb{R}^{2N}$ where $0 \leq \alpha \leq 1$. The true-scaled value of the beamforming action is calculated as follows:

$$\Re(p_{k,n}) = \frac{P_t}{\sqrt{\Upsilon_k}} \alpha_{2n-1}, \quad \Im(p_{k,n}) = \frac{P_t}{\sqrt{\Upsilon_k}} \alpha_{2n}, \tag{13}$$

where $\Upsilon_k = \sum_{n=1}^{n=2N} \alpha_n, \forall k \in \mathcal{K}$. Regarding the common rate action $\mathbf{c}$, the Softmax function is used to sum the K elements of $\bar{\mathbf{c}} = [\beta_1, \dots \beta_K]^T \in \mathbb{R}^K$ into 1. The true-scaled value of the common rate value is calculated as follows:

$$R_c = \min(R_1^c, \dots, R_K^c) \left(\sum_{k=1}^{K} \beta_k \right). \tag{14}$$

The velocity and direction angle of the UAV moving function are rescaled:

$$V_0 = \bar{V}_0 V_{max}, \quad \varphi_0 = \bar{\varphi}_0 2\pi. \tag{15}$$

Thus, the original action determined by the policy at each step is re-defined as

$$\bar{a}(t) = \{\alpha_1(t), \dots, \alpha_{2M}(t), \beta_1(t), \dots, \beta_K(t), \bar{V}_0(t), \bar{\varphi}_0(t)\}. \tag{16}$$

Due to the continuous action space, the Q-network is differentiable with respect to the action. Thus, the sampled policy gradient method is constructed to update the parameter sets of the policy, which is expressed as

$$\theta^{\mu} \leftarrow \theta^{\mu} + \frac{lr_{\mu}}{|B|} \sum_{s=1}^{B} \nabla_{\theta^{\mu}} \mu(s_s|\theta^{\mu}) \nabla_{a_s} Q(s_s, a_s|\theta^{Q})|_{a=\mu(s|\theta^{\mu})} \tag{17}$$

The weights of the two target networks are updated using a "soft" target update with constant $\daleth \ll 1$, expressed as

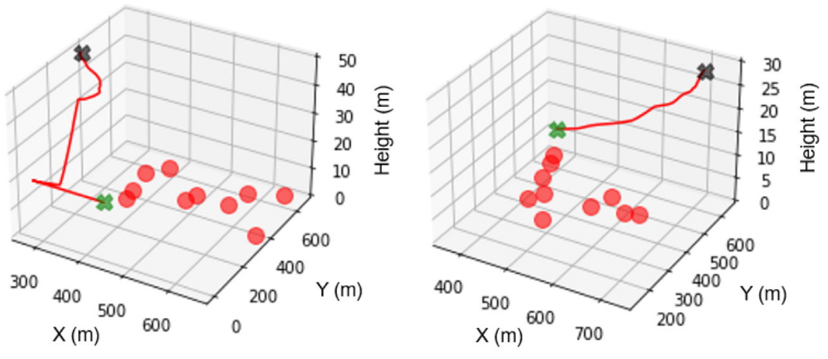
$$\theta \leftarrow \daleth\theta + (1 - \daleth)\theta. \tag{18}$$

The safe action-shaping DDPG (SAS-DDPG) algorithm for solving the maximum sum-rate problem in (8) is provided in Algorithm 1.

5. Simulation results

This section evaluates and compares the performance of the SAS-DDPG to other benchmark schemes in different system scenarios through computer simulations. In particular, we utilized PyTorch with Python 3.8.5 on a server with an Intel Core i7-12700 CPU, an Nvidia RTX 3070 GPU, and 32 GB of memory to carry out the simulations. The system parameters are summarized in Table 1.

First, the performance of the SAS-DDPG approach is evaluated in various scenarios. To explicitly emphasize the superiority of the combination of the SAS-DDPG approach and the RSMA technique, we simulate the proposed approach compared to the original DDPG approach in the NOMA and RSMA settings. In addition, the soft actor–critic method, greedy, random, and the SAS-DDPG with trajectory only are considered as benchmark schemes. Regarding the greedy and random schemes, the action value is quantized into discrete space and find the optimum action at each time step.



(a) In training process (b) In testing process

Fig. 3. Records of the uncrewed aerial vehicle trajectory training in 2500 episodes and testing in 100 episodes with $K = 10$ and $N = 25$.

Algorithm 1 Safe action-shaping deep deterministic policy gradient algorithm

```
Initialize hyperparameters: $\gamma, lrc, lra, B, D,$ ⌝
Initialize critic network $Q(s, a|\theta^Q)$ and actor network $\mu(s|\theta^\mu)$ with weights $\theta^Q$ and $\theta^\mu$
Initialize the target network $Q'$ and $\mu'$ with weights $\theta^{Q'} \leftarrow \theta^Q$ and $\theta^{\mu'} \leftarrow \theta^\mu$
for episode = $1 \ldots E$ do
    Generate locations of the UAV and $K$ users
    Observe the initial state $s(1)$
    for step = $1 \ldots T$ do
        Observe state $s(t)$
        Select and execute overall action $\bar{a}(t)$ according to (12)
        Shape the actions according to (13), (14), and (15).
        Observe reward $r(t)$ and next state $s(t+1)$
        Store experience $(s(t), a(t), r(t), s(t+1))$ in buffer $D$
        Uniformly sample a batch of $B$
        Update parameter $\theta^Q$ by minimizing the loss according to (10)
        Update parameter $\theta^\mu$ using the sampled policy gradient method according to (17)
        Update the target networks according to
    end for
end for
Return $\theta^{\mu*}$
```

Fig. 3 records the simulated UAV trajectory for the scenario of $K = 10$ and $N = 25$. In the training process, the UAV travels with an abnormal behavior because it is learning the dynamic environment and the moving behavior of the K users. When the learning process has been completed, the trained DNN model is collected and tested in an environment of 100 episodes. Fig. 3(b) indicates that the proposed algorithm can effectively learn the dynamic environment of RSMA downlink communication with user mobility. In addition, Fig. 4 depicts the convergence comparison between the SAS-DDPG and the original DDPG. Due to the proposed action shaping functions, the SAS-DDPG has stabler and better performance with an increase of 2.42% regarding the total reward.

Fig. 5 compares the SAS-DDPG RSMA in terms of the ASR with benchmark schemes. Under the scenario of $N = 16$ and $K = 18$, the transmit power of the UAV is altered from 0 dBm to 30 dBm. Combining it with the RSMA technique, the ASR with RSMA achieves 3.52%, 6.36%, 5.22% higher results compared to ASR with NOMA in the case of the SAS-DDPG, original DDPG, and SAC, respectively. In addition, it is apparent that the proposed approach outperforms other benchmark schemes regarding the ASR. Thus, the proposed

Table 1
Simulation parameters.

Parameter	Value
Network topology	$1000 \times 1000 \times 50$
Large scale pathloss L_{h_k}	$30 + 22log(d_k^{IU})$
UAV maximum velocity	20 m/s
Noise variance	-94 dBm/Hz
Rician factor κ	10
Minimum QoS R_{th}	1 bps
Carrier wavelength	0.5
User velocity	1.2 m/s
User moving angle	$\mathcal{CN}(\pi, 1)$
Target network update rate, ⌝	1e−3
Epsilon decay rate, ϵ	1e−3
Training/Testing episodes, E_{train}/E_{test}	2000/100
Steps per episode, S	300
Buffer capacity D	1e6
Batch size B	128
Network learning rate lr	1e−3
Target update weight ⌝	1e−3
Discount factor γ	0.9

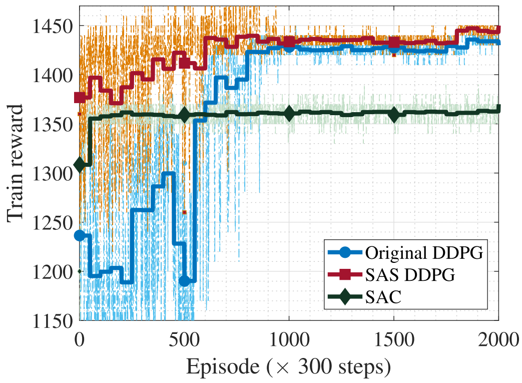


Fig. 4. Convergence performance comparison between safe action-shaping deep deterministic policy gradient, original deep deterministic policy gradient, and soft actor critic with $K = 10$ and $N = 25$.

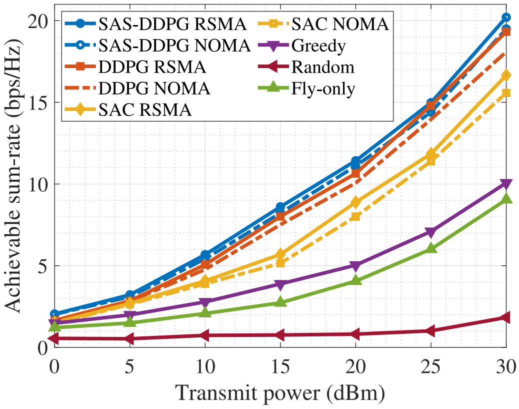


Fig. 5. Comparison of the achievable maximum sum rate versus the different transmit power values with $K = 18$ and $N = 16$.

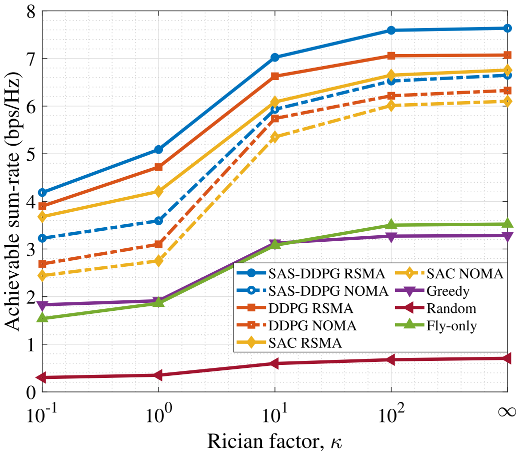


Fig. 6. Comparison of the achievable maximum sum rate versus the various fading environment factors with $K = 18$ and $N = 16$.

approach can be effectively applied to scenarios with multiple antennae and dense user deployment.

Fig. 6 illustrates the effect of the Rician factor of the SAS-DDPG RSMA scheme regarding the ASR compared to other benchmark schemes. In particular, the probability level of the LoS channel indicates by the κ value, which is changed to mimic different levels of uncertainty in channel conditions. For a scenario of $K = 10$ and $N = 16$, five cases of κ: $\{10^{-1}, 10^{0}, 10^{1}, 10^{2}, \infty\}$ are simulated. The SAS-DDPG combined with the RSMA technique outperforms the comparison schemes, with 7.38%, 11.52%, 57.02% and 90.76% higher results than the original DDPG, SAC, greedy, and random schemes. Thus, it is evident that the combination of the trajectory strategy, rate-splitting allocation, and beamforming design can enable efficient ASR maximization gains at different levels in a fading environment.

6. Conclusion

This study investigated the UAV-aided RSMA downlink communication system considering user mobility and URA antenna design. The sum-rate maximization problem was formulated by jointly optimizing the beamforming matrix, common rate allocation, and UAV trajectory design. The SAS-DDPG algorithm was proposed to address the formulated problem without requiring prior CSI information. Through the numerical simulation results, it is evident that the proposed algorithm efficiently infers dynamic CSI and determines instantaneous optimal solution. Nonetheless, there are many research challenges and issues, including propagation under imperfect CSI, efficient energy management, hardware constraint and heterogeneous IoT environment, which we should consider in the future work.

CRediT authorship contribution statement

Duc-Thien Hua: Conceptualization, Methodology, Software, Validation, Writing – review & editing, Visualization. **Quang Tuan Do:** Software. **Nhu-Ngoc Dao:** Supervision, Reviewing. **Sungrae Cho:** Supervision, Reviewing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Chung-Ang University, South Korea, Young Scientist Scholarship in 2021, and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00209125).

References

[1] N.-N. Dao, D.-N. Vu, W. Na, J. Kim, S. Cho, SGCO: Stabilized green crosshaul orchestration for dense IoT offloading services, IEEE J. Sel. Areas Commun. 36 (11) (2018) 2538–2548.
[2] D.T. Hua, Q.T. Do, T.V. Nguyen, C.M. Ho, S. Cho, Trajectory design in multi-UAV-assisted RSMA downlink communication, in: 2022 13th International Conference on Information and Communication Technology Convergence, ICTC, 2022, pp. 1048–1050.
[3] T.-H. Nguyen, T.P. Truong, N.-N. Dao, W. Na, H. Park, L. Park, Deep reinforcement learning-based partial task offloading in high altitude platform-aided vehicular networks, in: 2022 ICTC Conference, IEEE, 2022.
[4] J. Won, D.-Y. Kim, Y.-I. Park, J.-W. Lee, A survey on UAV placement and trajectory optimization in communication networks: From the perspective of air-to-ground channel models, ICT Express (2022).
[5] D.T. Hua, D.S. Lakew, S. Cho, DRL-based energy efficient communication coverage control in hierarchical HAP-lap network, in: 2022 International Conference on Information Networking, ICOIN, 2022, pp. 359–362.
[6] J.S. Yeom, Y. bin Kim, B.C. Jung, UAV-assisted cooperative downlink NOMA with virtual full-duplex operation, ICT Express 5 (4) (2019) 240–244.
[7] O. Abbasi, H. Yanikomeroglu, Transmission scheme, detection and power allocation for uplink user cooperation with NOMA and RSMA, IEEE Trans. Wireless Commun. 22 (1) (2023) 471–485.

[8] O. Dizdar, Y. Mao, B. Clerckx, Rate-splitting multiple access to mitigate the curse of mobility in (massive) MIMO networks, IEEE Trans. Commun. 69 (10) (2021) 6765–6780.

[9] Z. Yang, M. Chen, W. Saad, W. Xu, M. Shikh-Bahaei, Sum-rate maximization of uplink rate splitting multiple access (RSMA) communication, IEEE Trans. Mob. Comput. 21 (7) (2022) 2596–2609.

[10] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, H.V. Poor, Rate-splitting multiple access: Fundamentals, survey, and future research trends, IEEE Commun. Surv. Tutor. 24 (4) (2022) 2073–2126.

[11] A.A. Ahmad, J. Kakar, R.-J. Reifert, A. Sezgin, UAV-assisted C-RAN with rate splitting under base station breakdown scenarios, in: 2019 IEEE International Conference on Communications Workshops (ICC Workshops), 2019, pp. 1–6.

[12] W. Jaafar, S. Naser, S. Muhaidat, P.C. Sofotasios, H. Yanikomeroglu, On the downlink performance of RSMA-based UAV communications, IEEE Trans. Veh. Technol. 69 (12) (2020) 16258–16263.

[13] H. Bastami, M. Letafati, M. Moradikia, A. Abdelhadi, H. Behroozi, L. Hanzo, On the physical layer security of the cooperative rate-splitting-aided downlink in UAV networks, IEEE Trans. Inf. Forensics Secur. 16 (2021) 5018–5033.

[14] Deep reinforcement learning-based model-free path planning and collision avoidance for UAVs: A soft actor–critic with hindsight experience replay approach, ICT Express (2022).

[15] G. Geraci, A. Garcia-Rodriguez, M.M. Azari, A. Lozano, M. Mezzavilla, S. Chatzinotas, Y. Chen, S. Rangan, M.D. Renzo, What will the future of UAV cellular communications be? A flight from 5G to 6G, IEEE Commun. Surv. Tutor. 24 (3) (2022) 1304–1335.

[16] D. Tse, P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, USA, 2005.

[17] S.K. Yong, J. Thompson, Three-dimensional spatial fading correlation models for compact MIMO receivers, IEEE Trans. Wireless Commun. 4 (6) (2005) 2856–2869.

[18] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018, CoRR. arXiv:1801.01290.

[19] D.S. Lakew, V.D. Tuong, N.-N. Dao, S. Cho, Adaptive partial offloading and resource harmonization in wireless edge computing-assisted IoE networks, IEEE Trans. Netw. Sci. Eng. 9 (5) (2022) 3028–3044.