K. Guo, M. Wu, X. Li, H. Song and N. Kumar, "Deep Reinforcement Learning and NOMA-Based Multi-Objective RIS-Assisted IS-UAV-TNs: Trajectory Optimization and Beamforming Design," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2023.3267607.

https://doi.org/10.1109/TITS.2023.3267607

Deep Reinforcement Learning and NOMA-Based Multi-Objective RIS-Assisted IS-UAV-TNs: Trajectory Optimization and Beamforming Design

Kefeng Guo, Min Wu, Xingwang Li, *Senior Member, IEEE*, Houbing Song, *Fellow, IEEE*, and Neeraj Kumar, *Senior Member, IEEE*

***Abstract*— In this paper, we discuss the co-optimized performance of multi-reconfigurable intelligent surface (RIS)-assisted integrated satellite-unmanned aerial vehicle-terrestrial network (IS-UAV-TN), where the multiple vehicle users are applied to the network under consideration. The performance optimization of IS-UAV-TNs faces two major challenges: one is the obstacles in the transmission path and the other is the highly dynamic communication environment caused by the UAV movement for the multiple ground vehicle users. To tackle these above issues efficiently, we will install RIS on the UAV for the purpose of reshaping the wireless transmission path. In addition, non-orthogonal multiple access (NOMA) protocols are considered as a new paradigm to address spectrum shortage and enhance connection quality. Considering the UAV energy consumption, the satellite transmission beamforming matrix and RIS phase shift configuration, a multi-objective optimization problem is proposed to maximize the system achievable rate and minimize the UAV energy consumption during a specific mission. On this foundation, to facilitate the online decision problem, the deep reinforcement learning (DRL) algorithm is utilized to achieve real-time interaction with the communication environment. A multi-objective deep deterministic policy gradient (MO-DDPG) algorithm is proposed to search for sub-optimal solutions about the learning problem of multi-objective control policies in IS-UAV-TNs. Experimental results show that the method can simultaneously consider three optimization objectives and effectively adjust the optimal update policy according to the settings of different weight parameters.**

***Index Terms*— Deep reinforcement learning (DRL), reconfigurable intelligent surface (RIS), integrated satellite-unmanned aerial vehicle-terrestrial networks (IS-UAV-TNs), non-orthogonal multiple access (NOMA), multi-objective DDPG.**

I. Introduction

OWING to the rapid development of the Internet of Things (IoT) and the Internet of vehicles (IoV), the integrated satellite-terrestrial networks (ISTNs), which can provide heterogeneous services, seamless coverage and high data throughput for anyone and anything, have attracted widespread and significant attention as a reliable emerging candidate network architecture [1], [2], [3], [4].

Although integrating satellite networks into terrestrial networks has been proven to improve significantly, the severe path loss between terrestrial user equipment and Geostationary Orbit (GEO) satellites poses a major challenge due to transmission distances [5]. Thus, a communication relay is needed to amplify and forward the signal. The unmanned-aerial vehicle (UAV)-based relay communication is expected to be key technology due to its flexible capability and highly profitable benefits in ISTNs to achieve sustainable management, supervision and control of physical infrastructure.

Despite the evident merits of IS-UAV-TNs communications, this also causes serious concern about the limited spectrum resources and rapidly increased energy consumption [6], [7]. On this foundation, the power-domain non-orthogonal multiple access (NOMA) scheme can support a large number of multi-user access, especially, the application of NOMA to ambient backscatter communication technology is proposed in [8] and [9] as a reliable alternative to support large-scale heterogeneous services for 6G IoV networks. Based on the above analysis, the utilize of NOMA as a very promising

Manuscript received 1 December 2022; revised 23 February 2023 and 6 April 2023; accepted 10 April 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62001517; in part by the Henan Scientific and Technological Research Project under Grant 232102211073; in part by the Key Laboratory of Cognitive Radio and Information Processing, Ministry of Education (Guilin University of Electronic Technology) under Grant CRKL220203; and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515010999. The Associate Editor for this article was K. P. Tran. *(Corresponding author: Min Wu.)*

Kefeng Guo and Min Wu are with the School of Space Information, Space Engineering University, Beijing 101407, China (e-mail: guokefeng.cool@163.com; 1800022837@pku.edu.cn).

Xingwang Li is with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China (e-mail: lixingwangbupt@gmail.com).

Houbing Song is with the Department of Information Systems, University of Maryland Baltimore County (UMBC), Baltimore, MD 21250 USA (e-mail: h.song@ieee.org; songh@umbc.edu).

Neeraj Kumar is with the Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology (Deemed to be University), Patiala 147004, India, also with the School of Computer Science, University of Petroleum and Energy Studies, Dehradun 248007, India, also with the Department of Electrical and Computer Engineering, Lebanese American University, Beirut 1102 2801, Lebanon, also with the Computer Science and Engineering Department, Chandigarh University, Mohali 160012, India, and also with the Faculty of Computing and IT, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: nehra04@gmail.com; neeraj.kumar@thapar.edu).

Digital Object Identifier 10.1109/TITS.2023.3267607

access technology in IS-UAV-TNs, which has the potential to mitigate multipath and shadow fading [10], [11]. Among them, the problem of residual transceiver hardware defects in collaborative NOMA networks is very comprehensively explained in [12] and [13], while the outage probability (OP) and ergodic capacity (EC) are derived for cooperative and non-cooperative NOMA networks.

Apart from the limited spectrum resources, another challenge affecting the communications quality in IS-UAV-TNs is the instability of the transmission link, especially over cities with high density at low altitudes, and may encounter potential obstacles during UAV flight. To address this issue, reconfigurable intelligent surfaces (RISs) have been put forward as a new paradigm for intelligently changing wireless propagation environment [14]. RIS has numerous low-cost nearly passive reflective elements, every element regulated by pin-diodes or varactors, that is capable of constructively boosting the power of the received signal or destructively suppressing redundant interference by adjusting the phase shifts and/or amplitudes desired by intended ground vehicle users [15].

Since, the signal transmission process involves multiple optimization objectives, including the transmit beamforming weight vectors for the satellite, the design of phase shift for RIS, and the constraint of high-quality trajectory for UAV [16], [17]. As a result, the IS-UAV-TNs faces high-dimensional optimization problems that are challenging to solve using traditional methods, particularly when aiming to ensure the UAV's high-quality path planning to effectively avoid obstacles [18]. Traditional efficient solutions always require huge training overhead and need to meet the following requirements, such as a large amount of heterogeneous data generation, efficient data sensing, real-time data processing capabilities, and greater communication request arrival rates. Considering the ultimate goal of achieving harmonic co-existence among all heterogeneous wireless systems, the goal of developing new optimized intelligent algorithms was set.

Driven by the development of model-free artificial intelligence (AI) algorithm framework, such as the reinforcement learning (RL), deep learning (DL) and deep reinforcement learning (DRL), extensive industrial activities and operations are moving towards real-time automation and improvement. Among the existing AI approaches, DRL has emerged as an effective technique for handling explosive massive amounts of communication data, management of systems and resources mathematically, intractable nonlinear non-convex problems, even though the highly computational problem of studying and building knowledge networks about wireless channels without knowledge of channel models and the multiple terrestrial/vehicle users movement patterns [19]. And, we also finding optimal solutions to complex optimization problems by observing the reward after interacting with the wireless environment, thus enabling efficient algorithm design [20].

A. Related Works

Now, according to the above mentioned technologies, the related works can be discussed from the following aspects:

1) For the NOMA in IS-UAV-TNs Aspect: In the existing literature, a flurry of research works have been extensively studied on NOMA-based IS-UAV-TNs to harvest their benefits. In [21], the authors investigated the multi-objective optimization of uplink communication utilizing NOMA scheme in multiple UAV communication scenarios. Besides, the authors in [22] considered the UAV as an aerial relay to support two groups of ground users. Hence, with the explosion of demand for IoT and satellite communication services, the integration of NOMA into IS-UAV-TNs can significantly improve the utilization of frequency resource and has been proved to be a promising and effective method to achieve significant performance improvement of future wireless mode. In [23], the performance of NOMA-based IS-UAV-TNs with multi-objective optimization problem was discussed. A two-step solution was exploited to measure a max-min problem in terms of UAV's energy efficiency. The impacts of UAV as aerial base stations based on wireless powered communication (WPC) technology in NOMA-based IS-UAV-TNs were investigated in [24].

2) For the RIS-Assisted Transmission Aspect: Currently, a significant amount of work has been conducted to analyze the performance of networks assisted by RIS. The design of phase shifts, also known as passive beamforming, is crucial in fully utilizing the potential of RIS. For this reason, the issue of phase shift design has been extensively studied under different communication environment settings. The authors of [25] developed a covert communication scheme that utilizes an UAV equipped with an RIS to achieve maximum covert transmission rates. In [26], the authors proposed that the RIS-assisted air-ground networks was considered in two communication scenarios. In the first one, the RIS was mounted on a UAV to increase mobility and enable the creation of a direct line-of-sight link between the transmitter and receiver. The second scenario involved the use of an RIS to amplify the signal of the intended user while simultaneously suppressing signals from eavesdropping users. In [27], the authors investigated the OP and system average sum-rate in multi-RIS assisted communication system, and the closed-form expression of the asymptotic sum-rate are derived relying on the theory of extreme values. In [28], the authors investigated a four-stage optimization algorithm in multi-RIS-assisted multi-UAV mobile edge computing system. This algorithm simultaneously optimizes both UAV trajectories and RIS phase shifts. Besides, the authors in [29] proposed a model of a RIS-assisted vehicle communication system the performance under different signal-to-noise and interference scenarios was discussed. In [30], the authors proposed the secrecy issues in RIS-assisted satellite-ground relay networks with multiple UAV eavesdroppers, examining in detail the secrecy outage probability (SOP) through theoretical derivation and simulation experiments.

However, the above literatures have not taken into account the situation of the RIS-assisted wireless communications in IS-UAV-TNs, which will become an indispensable component of auxiliary communication due to the large coverage area and changing signal propagation paths. Hence, the consideration of RIS-assisted IS-UAV-TNs will be meaningful and have the ability to further improve communication efficiency.

3) For the RIS-Assisted NOMA Aspect: Recent studies have examined the potential profits of RIS-assisted NOMA networks [31], [32]. In [33], the authors investigated the performance of RIS-assisted NOMA networks under the circumstances of imperfect successive interference cancellation (ipSIC) and perfect successive interference cancellation (pSIC), and discussed the network throughput and energy efficiency of RIS-NOMA network under the mode of delayed limited and tolerated transmission. In a similar context, [34] identified that RIS is used to change the wireless communication environment of cell-edge users in a dual-cell NOMA network by adjusting the phase, and a method is given for minimizing the total transmit power while satisfying the signal to interference plus noise ratio (SINR) requirements. The RIS-NOMA scheme mentioned in the above the references did not take into account the division of users requirements, which is crucial for the implementation of RIS-NOMA scheme, as the computational complexity of the RIS-NOMA scheme increases with the number of service users [35].

4) For the DRL in Communication Aspect: Due to the successful application of the latest AI algorithms in numerous fields, more and more DRL algorithms are being used to tackle with communication network tasks such as wireless communication resource allocation, management of networked systems and resources [36], transmit power control, users requirement prediction, signal anomaly detection [37] and multi-objective optimization problems [38]. In [39], the authors considered using DDPG framework to jointly optimize the transmit power of the secondary transmitter and the RIS reflect beamforming in RIS-assisted cognitive radio system and compared results with traditional algorithm. The authors in [40] proposed an optimization objective to minimize UAV energy consumption through jointly the movement of the UAV, the RIS phase shift, the UAV's power allocation policy and dynamic decoding order in NOMA-based UAV systems. Additionally, the phase shift of the RIS can be aligned by adjusting the three-dimensional motion trajectory of the UAV and the real-time positioning of the ground terminal to achieve maximum system data transmission rate. In [41], the authors investigated the optimization of energy efficiency in RIS-assisted cellular networks driven by energy harvesting techniques. To solve this problem, a original framework on the basis of DRL algorithm was considered in which the base station received the state information, including user CSI feedback and the available energy disclosed by the RIS. From the above analysis, DRL is able to study and construct wireless channels by observing the feedback rewards from the surrounding communication environment, which leads to efficient algorithm framework design.

B. Motivation and Contributions

Considering that NOMA-based system performance depends on the difference between channel correlation and channel gain, the traditional channel model is determined by the propagation environment [42]. Thus, the channel interference environment becomes more random and complex, and the decoding sequence needs to be designed under various wireless channel conditions. The aforementioned factors lead to a highly coupled problem of optimizing the performance of UAV mobility, RIS configuration, and downlink beamforming, making it challenging to obtain the optimal solution using traditional iterative approaches [43].

To tackle with the above issues, we study a RIS-assisted IS-UAV-TNs NOMA-based downlink system, in which we install the RIS on the side of UAV to receive signals from satellite and then reflect them to ground vehicle users [44]. More specifically, with the assistance of a RIS-equipped UAV, the satellite simultaneously transmits signals to the VUs via the NOMA protocol to provide an aggregated virtual line-of-sight (LoS) link. This proposed framework introduces a new flexible paradigm in efficient spectrum sharing between satellite and ground multi-vehicle users. Our major efforts can be concluded below.

- Firstly, we propose a novel framework for RIS-assisted IS-UAV-TNs communication architecture, which adopts the NOMA technology for promoting a flexible multiple access. On this foundation, through joint improvement of UAV trajectory, RIS configuration and downlink emission beam formation, the energy efficiency maximization problem of the system is formulated to ensure the full capability and minimum energy requirements of UAV and terrestrial vehicle users (VUs).
- Secondly, the classical DDPG algorithm framework of one-dimensional reward is extended to multi-dimensional reward and put forward a distributed robust DRL algorithm on the basis of the multi-objective DDPG structure. The RIS passive phase shift and transmit beamforming are aligned through an online UAV trajectory learning to achieve the alignment of various signals for the highest data transfer rates. Thus, by adjusting the RIS passive phase shift, the transmission path can be changed to achieve more efficient signal redirection, thus achieving better propagation conditions in IS-UAV-TNs.
- Finally, the effectiveness of the proposed scheme is verified by simulation experiments, and the flexibility of MO-DDPG algorithm in optimization strategy is proved to be better than that of traditional rule-based strategy. By updating the network weight parameters through a soft update policy, the optimal policy can be modified which optimized the multiple objectives problem collaboratively under different priority levels.

The remaining of this paper is arranged as follows. In Section II, we describe the proposed system model and mathmatize our system transmission sum rate maximization problem. In Section III, we focus on the DRL-based algorithm for the jointly optimization problem, variables to be optimized include UAV trajectory, the RIS phase shift and transmit beamformingn. Finally, simulation results and analysis are elaborated in Section IV to verify the performance from different performance indicators of the proposed MO-DDPG algorithms framework, whereas Section V give conclusions about this paper. In addition, the main notations mentioned in this paper are shown in Table I.

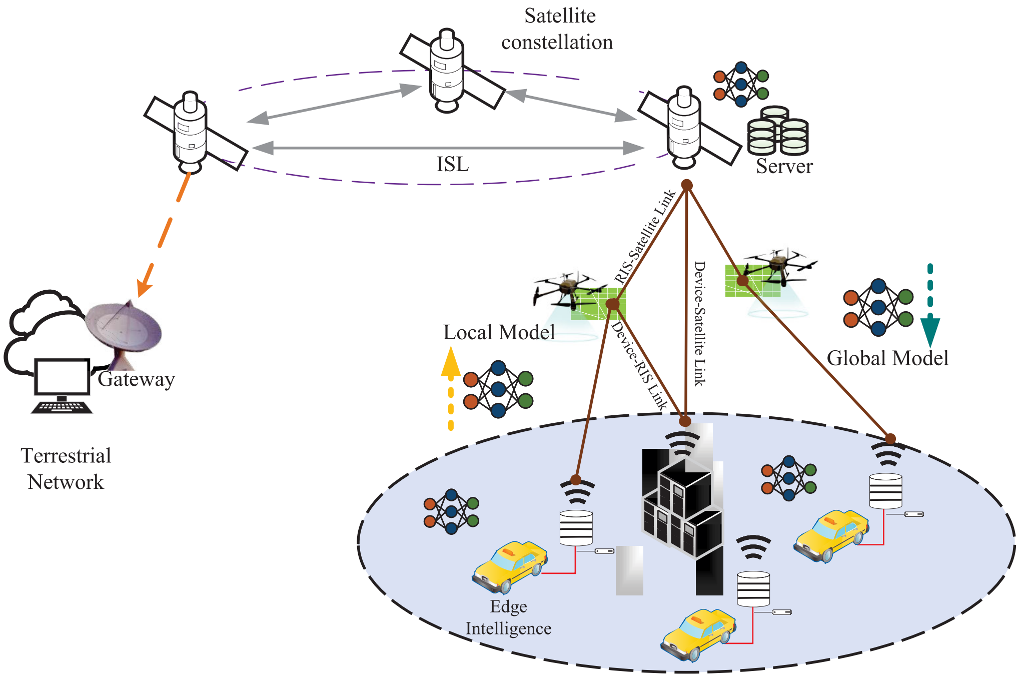


Fig. 1. Illustration of NOMA in multi-UAV-RIS-assisted satellite-vehicle networks.

TABLE I

LIST OF NOTATIONS

Notations	Description
N	Number of UAV or RIS (A UAV equipped a RIS)
K	Number of vehicle users
M	Number of RIS reflective elements
P_{max}	Satellite transmit maximum power
p_k	Transmit power of satellite to k-th VUs
$\mathbf{w}_k$	Transmit beamforming matrix of satellite to k-th VUs
$\mathbf{G}_{S,k}$	The channel from satellite to k-th VUs
$\mathbf{h}_{R,k}$	The channel from RIS to k-th VUs
$\mathbf{G}_{SR}$	The channel from satellite to RIS
θ^m	The m-th reflecting element's phase shift
E_k	The total energy consumption of UAV
$T_{\max}$	UAV maximum flight time
$q_{S,n}^t$	The n-th UAV initial position
$q_{D,n}^t$	The n-th UAV final position
x_k^t, y_k^t	Coordinate of k-th VUs
R_k	Instantaneous system achievable rate
σ^2	Power of white Gaussian noise
$a^{(t)}$	Action in MO-DDPG algorithm
$s^{(t)}$	State in MO-DDPG algorithm
$R^{(t)}$	Raward in MO-DDPG algorithm

Other Notations: In this paper, for general channel representation $\mathbf{G}$, $\mathbf{G}(i, j)$ is the entry at the i-th row and the j-th column. And, $(\centerdot)^H$ and $(\centerdot)^T$ denote the conjugate and transpose operation of channel matrix. Moreover, $\mathbb{C}^{M\times N}$ denotes the set of $M \times N$ complex vectors, $\mathbb{E}[\centerdot]$ denotes the expectation operation.

II. System Model and Problem Formulation

A. System Model Description

As can be seen from Fig. 1, we consider the downlink RIS-assisted IS-UAV-TNs communications system model, which consists of a satellite, N UAVs equipped with N RISs (each UAV is equipped with one RIS) and multiple terrestrial vehicle users (VUs) to be served. It is supposed that the satellite configured with a single omni-directional antenna is applied to provide communication service for a total K single antenna terrestrial vehicle users. Moreover, we consider that all VUs are roaming randomly in a fixed area. In an effort to create a tandem virtual LoS propagation path using the UAV flexibility between the satellite and the vehicle users by passively reconfiguring the incident signal at the receiver using RIS, so we can install the RIS on the side of the UAV, and each RIS is made up of M reflective elements, thus enhancing the service quality of wireless communcation.

The whole communication includes two links, the signal transmission direct link and the signal transmission reflected link. The direct link is the satellite that transmits signals directly to VUs without being relayed or reflected. The reflected link is the signal that first transmitted to the reflecting RIS in UAV and then reflected by the RIS to VUs. The UAV is only used as an airborne RIS-mounted mobile platform in our paper and is not involved in communications. By introducing the UAV, the RIS deployment scheme is more flexible, thus improving signal transmission efficiency. To distinguish different channel representations, we simply define the channel model vector coefficients from the satellite to the k-th terrestrial VUs, the RIS to the k-th terrestrial VUs, the satellite to the RIS as $\mathbf{G}_{S,K}^H \in \mathbb{C}^{1\times M}$, $\mathbf{h}_{R,K}^H \in \mathbb{C}^{1\times M}$, and $\mathbf{G}_{SR} \in \mathbb{C}^{1\times 1}$, respectively.

The diagonal phase shift matrix of the i-th RIS is denoted by $\Theta_i = diag\left(e^{j\theta_i^1}, e^{j\theta_i^2}, \ldots, e^{j\theta_i^M}\right)$, where $\theta_i^m \in [0, 2\pi)$ represents the phase shift of the m-th reflecting element at the i-th RIS with $M = M_R \times N_R$. And, the actual discrete phase-shift values are need to be considered, i.e. $\theta_n^m \in \{1, \Delta\theta, \ldots, (L-1)\Delta\theta\}$, where $\Delta\theta_i^m = 2\pi/L_i$ indicates the number of i-th RIS discrete phase shift levels. Thus, a joint optimization study on RIS passive beamforming and satellite active transmission beamforming is carried out on the basis of considering UAV motion trajectory. The key point is to investigate the performance of RIS-assisted IS-UAV-TNs under the NOMA access scheme using DRL framework. Hence, we assume that channel that channel large-scale fading has been compensated and the perfect CSI is available. This can be achieved by utilizing response and training data feedback from signal transmitters, as has already been implemented in DVB-S2 [45]. In addition, through the feedback and training from terrestrial VUs, the assumed perfect CSI is shared during transmission.

B. Overview of Channel Model

As far as the downlink channel model between the satellite and terrestrial VUs is concerned, we should take into account the free space transmission loss, the rain attenuation and the beam gains. Thus, it can be expressed as

$$\mathbf{G} = \sqrt{G_{s,t} G_{s,r} C_s} \odot \xi^{-\frac{1}{2}} \odot \mathbf{b}^{\frac{1}{2}} \odot e^{-j\phi}, \tag{1}$$

where, $G_{s,t}$ is the satellite launch gain, $G_{s,r}$ is the parabolic antenna gain for the satellite service user. Referring to the ITU recommendations, $G_{s,r}$ can expressed as

$$G_{s,r} = \begin{cases} G_{s,r,\max} - 2.5 \times 10^{-3} \left(\dfrac{d_s^r \theta_s}{\lambda} \right)^2, & 0^\circ < \theta_s < \theta_a \\ 2 + 15 \log \dfrac{d_s^r}{\lambda}, & \theta_a < \theta_s < \theta_b \\ 32 - 25 \log \theta_s, & \theta_b < \theta_s < 48^\circ \\ -10, & 48^\circ < \theta_s < 180^\circ, \end{cases} \tag{2}$$

where $G_{s,r,\max}$ denotes the maximum receive gain in the axial direction of the parabolic antenna, θ_s is the off-axis angle of the user relative to the satellite [46], d_s^r is the antenna diameter, and $\theta_a = \frac{20\lambda}{d_s^r} \sqrt{G_{s,r}^{\max} - \left(2 + 15 \log \frac{d_s^r}{\lambda}\right)}$ and $\theta_b = 15.85 \left(\frac{d_s^r}{\lambda}\right)^{-0.6}$ represent the angular value, respectively. $C_s = (\lambda / 4\pi d_s)^2$ denotes the free space loss, where λ is the carrier wavelength and d_s represents the distance between the satellite and the intended signal receiver, ξ denotes the rain attenuation coefficient which satisfies $\ln\left(\xi_k^{\mathrm{dB}}\right) \sim \mathcal{CN}\left(\mu_k, \sigma_k^2\right)$ with μ_k and σ_k being mean value and variance which are related to the satellite communication frequency, polarization mode and the location of the served user, respectively. And, the k denotes the number of satellite array antennas, $\mathbf{b}$ represents the beam gain matrix, by taking into account the far-field characteristic of satellite communication, every column vector of $\mathbf{b}$ can be found in Eq.(3) of [47].

Although considering that the actual flight altitude of UAV in the communication environment is higher than most buildings, it is assumed that the channel model between RIS and terrestrial VUs follows the Rician fading channel model, which indicates that there is still exists LoS transmission link between them. Thus, the transmission link from the RIS to VUs includes the LoS link and the NLoS link, which can be expressed as

$$\mathbf{h}_{RE} = \sqrt{\frac{K_{RE}}{K_{RE}+1}} \mathbf{h}_{LOS} + \sqrt{\frac{1}{K_{RE}+1}} \mathbf{h}_{NLOS}, \tag{3}$$

where K_{RE} is the Rician coefficient of the RIS-VUs link, $\mathbf{h}_{LOS}$ and $\mathbf{h}_{NLOS}$ are the LoS component and the NLoS component, respectively. In fact, $\mathbf{h}_{LOS}$ depends primarily on where the UAV is located, which can be modeled as [48]

$$\mathbf{h}_{LOS} = \sqrt{G_e C_e} \mathrm{vec}\left(\mathbf{A}\left(\theta_r, \varphi_r\right)\right) \tag{4}$$

where G_e is the VUs receive gain, C_e denotes the free space loss between RIS and VUs, and $\mathbf{A}(\theta_r, \varphi_r)$ is the RIS-VUs channel matrix which can expressed as

$$\mathbf{A}\left(\theta_r, \varphi_r\right) = \mathbf{a}_x\left(\theta_r, \varphi_r\right) \mathbf{a}_y^H\left(\theta_r, \varphi_r\right), \tag{5}$$

where θ_r and φ_r are the signal departure pitch and departure azimuth of the VUs relative to the RIS, respectively. And, $\mathbf{a}_x(\theta_r, \varphi_r)$ and $\mathbf{a}_y^H(\theta_r, \varphi_r)$ are the x- and y-axis guidance vectors of the RIS, respectively.

$$\mathbf{a}_x\left(\theta_\mathrm{r}, \varphi_\mathrm{r}\right) = \left[1, \mathrm{e}^{\mathrm{j}\frac{2\pi d_x}{\lambda}\sin\theta_\mathrm{r}\cos\varphi_\mathrm{r}}, \cdots, \mathrm{e}^{\mathrm{j}\frac{2\pi d_x}{\lambda}(M_R-1)\sin\theta_\mathrm{r}\cos\varphi_\mathrm{r}}\right]^\mathrm{T}, \tag{6}$$

$$\mathbf{a}_y\left(\theta_\mathrm{r}, \varphi_\mathrm{r}\right) = \left[1, \mathrm{e}^{\mathrm{j}\frac{2\pi d_y}{\lambda}\sin\theta_\mathrm{r}\sin\varphi_\mathrm{r}}, \cdots, \mathrm{e}^{\mathrm{j}\frac{2\pi d_y}{\lambda}(N_R-1)\sin\theta_\mathrm{r}\sin\varphi_\mathrm{r}}\right]^\mathrm{T}, \tag{7}$$

where d_x and d_y is expressed as the physical spacing of adjacent reflection elements in the horizontal and vertical directions of RIS. And, M_R and N_R denotes the number of reflecting elements in the vertical and horizontal directions of the RIS, respectively.

C. Signal Transmission Model for Orthogonal Multiple Access Scheme

OMA allows each user to completely separate unwanted signals from the required signals by allocating different blocks of orthogonal resources to each user. By analyzing the characteristics of the channel model for the above signal transmission process and considering the linear transmit precoding on the satellite, the signal transmitted by satellite with the orthogonal multiple access (OMA) scheme can be expressed as

$$\mathbf{x} = \sum_{k=1}^{K} \sqrt{p_k} \mathbf{w}_k s_k, \tag{8}$$

where denotes the signal power transmitted from the satellite to k-th VUs, $\mathbf{w}_k$ represents the transmit beamforming vector applied at the satellite, and s_k denotes the transmitted data symbols for k-th VUs [49]. The signal received at k-th VUs can be given in the following

$$y_k^{OMA} = \left(\mathbf{G}_{S,k}^H + \mathbf{h}_{R,k}^H \Theta \mathbf{G}_{SR}\right) \sum_{j=1}^{K} \sqrt{p_k} \mathbf{w}_k s_k + n_k, \tag{9}$$

where n_k denotes additive white Gaussian noise (AWGN) at the k-th VUs distributed as $n_k \sim \mathcal{CN}\left(0, \sigma_k^2\right)$. Based on

Eq. (8), the received SINR of the k-th VUs can be calculated as

$$\gamma_k^{OMA} = \frac{p_k\left|\left[\mathbf{G}_{S,k}^H + \mathbf{h}_{R,k}^H\Theta\mathbf{G}_{SR}\right]\mathbf{w}_k\right|^2}{\sum_{j\neq k}^{K} p_j\left|\left[\mathbf{G}_{S,j}^H + \mathbf{h}_{R,j}^H\Theta\mathbf{G}_{SR}\right]\mathbf{w}_j\right|^2 + \sigma_k^2}, \tag{10}$$

To eliminate mutual interference among multiple VUs, a linear precoding scheme based on zero-forcing (ZF) was always used on the satellite to determine transmit beamforming design [50]. In a departure from previous work, we optimize the active transmission precoding matrix using the DRL algorithm. Thus, the instantaneous downlink system achievable rate at the k-th VUs with the OMA scheme can be given by

$$R_k^{OMA} = \log_2\left(1+\gamma_k\right), \tag{11a}$$

$$= \log_2\left(1 + \frac{p_k\left|\left(\mathbf{G}_{S,k}^H + \mathbf{h}_{R,k}^H\Theta\mathbf{G}_{SR}\right)\mathbf{w}_k\right|^2}{\sum_{j\neq k}^{K} p_j\left|\left(\mathbf{G}_{S,j}^H + \mathbf{h}_{R,j}^H\Theta\mathbf{G}_{SR}\right)\mathbf{w}_j\right|^2 + \sigma_k^2}\right). \tag{11b}$$

D. Signal Transmission Model for NOMA Scheme

The other side of the signal transmission shield, the traditional OMA schemes is difficult to realize the trade-off between network throughput and user fairness, which will struggle to meet the exploding demand of large-scale terrestrial VUs. Therefore, the NOMA technology can distribute the same time/frequency/code to multiple users, which will be used in future terrestrial vehicle networks to improve spectrum efficiency. The NOMA-downlink decoding criterion is to preferentially decode the vehicle users with the worst channel quality. To be specific, the downlink NOMA-based system will allocate more transmission power to vehicle users with weaker channel quality, while less transmission power to VUs with stronger channel quality [51]. To simplify the expression, we set the synthetic channel coefficients experienced by the i-th VUs given by $\mathbf{h}_k = \mathbf{G}_{S,k}^H + \mathbf{h}_{R,k}^H\Theta\mathbf{G}_{SR}$. Thus, the received signal at the i-th VUs is given by

$$y_i^{NOMA} = \mathbf{h}_i\sqrt{p_i}\mathbf{w}_i s_i + \sum\nolimits_{j\neq i}\mathbf{h}_j\sqrt{p_j}\mathbf{w}_j s_j + n_i, \quad \forall j,\ \forall i, \tag{12}$$

where p_i means the downlink power assigned to the i-th VUs, and n_i denotes AWGN at the i-th VUs distributed as $n_i \sim \mathcal{CN}\left(0, \sigma_i^2\right)$. To alleviate the interference among multiple VUs, we adopt the successive interference cancellation (SIC) technique. At this point without loss of optimality, the channel coefficients of all ground VUs are calculated as $|\mathbf{h}_K| \leq \ldots \leq |\mathbf{h}_2| \leq |\mathbf{h}_1|$ [52]. Then, the transmission power of the satellite is then limited to the following constraints for a successful SIC implementation:

$$\Delta_i = p_i|\mathbf{h}_{i-1}|^2 - \sum_{j=1}^{i-1} p_j|\mathbf{h}_{i-1}|^2 \geq \rho_{\min}, \quad \forall i \geq 2, \tag{13}$$

where $\rho_{\min} > 0$ is used to differentiate the gaps in the decoded signals. When the above power and channel coefficients constraints are satisfied, the instantaneous downlink system achievable rate at the i-th VUs with the NOMA scheme, which can be obtained by

$$R_i = B_i\log_2\left(1 + \frac{|\mathbf{h}_i\mathbf{w}_i|^2 p_i}{|\mathbf{h}_j\mathbf{w}_j|^2\sum_{j=1}^{j\neq i} p_j + \sigma_i^2}\right), \quad \forall j \neq i,\ i \in K, \tag{14}$$

where B_i means the bandwidth allocated by the satellite to the i-th VUs.

E. Energy Dissipation Model for the UAV

Currently, according to different modes of communication maintenance, there exist two functional types of UAVs: rotary-wing UAVs and fixed-wing UAVs. In this paper, only fixed-wing UAVs are considered, and this type of UAV only performs communication services at the same altitude, without considering the vertical movement of the UAV. This is because the main goal of this paper is to develop a framework for UAV trajectory design and active-passive beamforming design using the advanced DRL framework. Once the proposed framework is adequately trained, it can be easily extended network input parameters to the three-dimensional trajectories, vehicle user distribution, UAV distribution, and other influencing factors. To promote the trajectory design, the total flying time T_K is dispersed into time slots with equal time intervals. Next, the horizontal position of the n-th UAV at time slot t can be represented by $q_n^t \in Q$, where $q \triangleq \{1, 2, .., Q\}$, $t \in T_k = \{1, 2, \ldots, T_K\}$ and T_K means the total number of transmission time slots. It is noted that $q_{S,n}$ and $q_{D,n}$ are defined as the centers of the initial and finial locations of the UAV's determined beforehand in fixed service area. Then, the horizontal trajectory without considering the vertical height of the UAV can be approximated as $\left\{q_{S,n}^t, q_2^t, \ldots, q_n^t, \ldots, q_{N-1}^t, q_{D,n}^t\right\}$ within a given deadline T_{max}. For vertical dimension, assume that UAV drones operate at fixed altitude $\mathbf{H}_0$ to provide long-term stable communication services.

Inspired by the rotary-wing UAV propulsion energy loss model, we define the blade profile power of the considered UAV as $E_1(t) = \frac{\delta}{8}\rho_a f A\Omega^3 R^3 + n_E(t)$ in hovering state, f indicating airframe drag ratio and rotor firmness, ρ_a denoting the air density, δ represents cross section drag coefficient, A implies the total area of the propeller, while Ω indicates the blade angular velocity and R is the radius of the whole rotor blade. Meanwhile, we also consider the slight difference in service type between the rotary-wing UAV and the considered UAV, and we set $n_E(t)$ as the error term as a supplement to this model.

The total consumed energy required by the UAV includes two major parts, namely the energy consumption associated with communication and propulsion movements. The first part is mainly used for receiving signals, signal processing, hardware computing equipment power consumption, etc., while the other part is mainly used to support the UAV hovering and power consumption during movement. It is noteworthy that the energy consumption associated with propulsion movement accounts for more than 95% of the total energy consumption. Since the movement of UAV requires much more energy consumption than that required for communication, so this paper mainly studies the energy consumption of UAV in movement or hovering. As a result, the total energy consumed by the k-th UAV carrying the RIS for communications services is as follows

$$E_k = E_1 T_k v + E_2 T_k + \varepsilon(t), \quad \forall k, \tag{15}$$

where E_1 and E_2 are two constants associated with the mechanical motion output power and the signal transmission loss, respectively. Moreover, the $\varepsilon(t)$ is a set of errors generated over the time whose value depends on the exact UAV movement pattern. The engine can be turned off when the UAV is in hover, which means that no energy is dissipated in this status.

F. Problem Formulation

In the NOMA-based RIS-assisted IS-UAV-TNs communication system, even if the original channel is not aligned, a single spatial direction can be used to serve multiple VUs, thus facilitating the implementation of NOMA. Because NOMA transmits multiple users' superimposed signals in the same time/frequency resource block, as well as the multi-antenna system itself will have serious inter-user interference, it is particularly important to introduce beamforming technology which is widely used in the system to reduce interference.

Thus, the purpose of this paper aims to jointly optimize the active transmit beamforming of the satellite, the passive reflection beamforming of RIS and the two-dimensional trajectory of the UAV for maximizing the system sum rate in the case of the UAV's minimum energy consumption [53]. Denote the $\mathbf{w} = [w_1, w_2, .., w_K]$ denotes the active transmit beamforming strategy, $\mathbf{Q} = \left[\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\right]^T$ represents the UAV served trajectory design scheme. Considering the influence of factors such as transmission power, RIS phase shifts and maneuverability of UAV, this paper gives the expression about the long-term optimal solution

$$\max_{\Theta,\mathbf{w},\mathbf{Q}} \frac{1}{T_k} \sum\nolimits_{t=1}^{T_k} \sum\nolimits_{k=1}^{K} \frac{R_k^t}{E_k} \tag{16a}$$

$$\text{s.t. } q_{S,n}^t = q_1^t, \quad q_N^{T_k} = q_{D,n}^t, \ \forall k, \tag{16b}$$

$$\Delta_k^t \geq \rho_{\min}, \quad p_k^t > 0, \ \forall k, \forall t, \tag{16c}$$

$$\sum\nolimits_{k=1}^{K} p_k^t < P_{\max}, \quad \forall k, \forall t, \tag{16d}$$

$$x_{\min} \leq x_k^t \leq x_{\max}, \tag{16e}$$

$$y_{\min} \leq y_k^t \leq y_{\max}, \tag{16f}$$

$$\left|\exp\left(j\theta_n^m\right)\right| = 1, \quad \forall n, \ \forall m. \tag{16g}$$

where Eq. (16a) denotes the optimization objective, the maximization of the optimization objective depends on maximizing the system achievable rate while minimizing the energy consumed by the UAV. This can be easily concluded that so as to maximize the system achievable rate, the UAV should be able to access more ground VUs during the mission time, increasing the number of channel transmissions to increase the total data transmission rate. Considering that UAV communication calculations consume less energy, mainly that generated by movement, hovering over the need for communication devices is the best option from this perspective. Eq. (16b) indicates the trajectory design of UAV, and specifies the starting and ending position to determine maximum service distance. Eq. (16c) and (16d) is to maintain the transmission power of the satellite and the power allocation constraint. The range of mission execution activities supported by the UAV energy is restricted between (16e) and (16f). In addition, Eq. (16g) indicates the need to satisfy the modulation constraint of the RIS.

It can be said that the formulated optimization problem Eq. (16) proposed in this paper to be optimized is a non-convex problem, mainly for the following reasons [54]. Firstly, multiple variables to be optimized, $(\Theta, \mathbf{w}, \mathbf{Q})$ are recursively related and tightly coupled in the objective optimization function. Secondly, owing to the RIS discrete phase shifts and the location-dependent channel model vector coefficients associated with the UAV position, the implementable rate R_k^t of the system is not a continuous function. Thirdly, the simultaneous movement of multiple UAVs also results in problem (15) difficult to address, especially in large-scale networks where even considering only the sub-problem of trajectory design is not possible. On the contrary, traditional algorithms are usually formulated to seek suboptimal solutions, maximizing the objective functions with alternating optimization techniques. At each training iteration, the sub-optimal $\mathbf{w}$ is solved by first modifying Θ while sub-optimal parameters Θ is derived by fixing the $\mathbf{p}$ and the UAV trajectory design strategy $\mathbf{Q}$ until the converge of algorithms. In this paper, instead of solving the challenging optimization problems with a mathematical method directly, we draw up the optimization problems in the framework of the advanced multi-objective DDPG (MO-DDPG) algorithm framework to obtain feasible $\mathbf{w}$ and Θ for improving energy efficiency.

III. MO-DDPG For UAV Trajectory Design, Transmit Beamforming Control, and RIS Configuration

In this section, the joint optimization problem of UAV trajectory design, transmit beamforming, and RIS configuration to maximize system transmission efficiency can be modeled as a Markov decision process (MDP), which is a common model for formulating such environment-interactive systems. Then, we design a highly sampling efficiency MO-DDPG algorithm framework to achieve the maximum expected long-term gain in the envisaged wireless environment.

A. MDP Formulation

The optimization multi-objective problem (16) can be modeled as a continuous decision process in the time horizon, that is, the next single time step is decided on basis of the current environmental interaction. In this context, the purpose of MDP is to find the optimal transmit beamforming strategy.

In this section, the MDP game can be expressed by a transition tuple with five elements $\langle S, A, R, T, \gamma\rangle$, standing for the state space, action space, reward function, transition policies and the discount factor, respectively.

The goal of learning agent is to find the optimal decision-making policy $\pi^*(s, a)$ by maximizing the state-value function $Q_\pi(s)$. Formally, the policy is a mapping from the state to the probability of choosing each possible action. If the agent follows the policy at t-th time step, then $\pi(a \mid s)$ is the possibility when current state $S_t = s$ takes action $A_t = a$.

The state-value function $Q_\pi(s)$ is defined as the expected cumulative discounted reward, which can be expressed as

$$\begin{aligned} Q_\pi(s) &= \mathbb{E}_\pi\left[G_t \mid S_t = s\right] \\ &= \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1} \mid S_t = s\right], \quad \forall s \in S, \end{aligned} \tag{17}$$

where, $\mathbb{E}_\pi$ denotes the expectation of the random variable when the agent follows the transition policy in the t-th time step, $\gamma \in [0, 1]$ indicates the discount factor, in which $\gamma \to 1$ indicates that long-term rewards are weighted more heavily than short-term rewards, $\gamma \to 0$ represents the opposite aspect. And, $G_t = \sum\limits_{k=0}^{\infty}\gamma^k R_{t+k+1}$ can be expressed as an action to maximize the sum of discounted reward obtained in the process.

Similarly, the value of the state s under the policy of taking actions a to obtain the expected reward is defined, which denotes as Q-value function $Q(s, a)$ for policy $\pi(s, a)$, and satisfies the Bellman equation. It can be shown as

$$\begin{aligned} Q_\pi(s, a) &= \mathbb{E}_\pi\left[G_t \mid S_t = s, A_t = a\right] \\ &= \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1} \mid S_t = s, A_t = a\right], \quad \forall a \in A. \end{aligned} \tag{18}$$

Meanwhile, we can further obtain the Bellman expectation equation, which can be expressed as

$$\begin{aligned} V_\pi(s) &= \mathbb{E}_\pi\left[G_t \mid S_t = s\right] \\ &= \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1} \mid S_t = s\right] \\ &= \sum_{a\in A}\pi(a \mid s)\left(R_s^a + \gamma\sum_{s'\in S}P_{ss'}^a Q_\pi(s')\right), \end{aligned} \tag{19}$$

where $R_s^a = \mathbb{E}\left[R_{t+1}|S_t = s, A_t = a\right]$ represents the reward function, and $P_{ss'}^a$ denotes the state transition probability matrix with $P_{ss'}^a = \mathbb{P}\left[S_{t+1} = s'|S_t = s, A_t = a\right]$. Note that, we consider the existence of a central controller in the formulated MDP model who is an agent for exploring the unknown wireless environment.

The expression and construction MDP in proposed DRL algorithm are described in detail below.

1) State Space: The state space information of DRL represents the information about the environment that the agent perceives and the changes brought about by its own actions. The state space information is the basis for the agent's decision making and assessment of its long-term benefits, and the design of the state directly determines the convergence, speed of convergence and ultimate performance of the DRL algorithm, a matter of great importance. Active transmit beamforming and RIS passive beamforming are designing depends on rewards during frequent state interactions between the agent and the wireless environment, which can consume large radio communication resources, significantly reducing the efficiency of the efficiency of successful system communications. Thus, in the t-th time step, we define that the state space $s^{(t)}$ consists of four components: the transmission power, the UAV's location q_n^t, the action from the last time step and the channel model matrix $\mathbf{h}_k$ and the $\mathbf{G}_r$. These state space elements are used to obtain good overall environmental information by configuring the RIS and transmit beamforming strategies and thus adjusting the real-time UAV trajectory. In addition, by randomly initializing the corresponding CSI, the missing channel information, which is common in IS-UAV-TNs, is overcome. The state space can be expressed as

$$s^{(t)} = \left\{p_k, q_n[t], a[t-1], \mathbf{h}_k[t], \mathbf{G}_r[t]\right\}. \tag{20}$$

Once the UAV location is determined, the active beamforming and passive beamforming are optimized based on the algorithmic reward as well as the current state, thus further guiding the adjustment of the UAV position. This alternating optimization algorithm challenges the available service time for uneven communication between balloons and satellites to include more target users in their accessible service areas. Combine the initial position $q_{S,n}$ and final position $q_{D,n}$ of UAV, avoid UAV flying out of the designated area, resulting in unnecessary energy consumption.

2) Action Space: The action space must, firstly, provide the possibility of achieving the desired goals, avoiding "state space blindness" in the task solution space that is out of reach and, in particular, ensuring adequate accessibility to high performance areas. Secondly, the action space should be as simple and efficient as possible in order to effectively reduce training difficulties and improve algorithm performance. On the one hand, the continuous action space can be reduced to zero and transformed into a discrete action space under the premise of satisfying the basic control accuracy, which can significantly compress the dimension of the solution space and improve the exploration efficiency; on the other hand, according to the actual situation, some basic actions can be organically combined to form macro actions. The developed MDP's action space consists of three main components, specifically the UAV's forward movement direction, the each reflecting element, as well as the satellite active transmission beam formation. Considering that the RIS passive reflection component includes both real and imaginary parts, the proposed framework has a mixture action spaces including discrete and continuous, making the proposed MDP problem extraordinary [55]. To meet this above challenge, the UAV maneuver direction and the satellite transmission power need to be discreted. Furthermore, the transformed action space converted only three components: 1) the UAV maneuver movement direction with leftward, forward, rightward, and backward, respectively. 2) the m-th element phase shift in

n-th RIS, i.e. $\theta_n^m \in \{1, \Delta\theta, \ldots, (L-1)\Delta\theta\}$; and 3) the transmit beamforming matrix for the satellite, i.e, $\mathbf{w}[t] = [w_1^t, w_2^t, \ldots, w_K^t]^T$. Thus, the action space can be shown as

$$a^{(t)} = \left\{q_n^t, \theta_n^{m,t}, \mathbf{w}[t]\right\}, \quad \forall t, \ \forall m, \ \forall n. \tag{21}$$

3) Reward: The design of the reward function is an extremely important aspect of DRL applications. By specifying and numerising the task objective, the reward acts as a special language for efficient communication between the optimization objective and the algorithm. As shown in the optimization problem (15), the goal of the joint optimization issues about UAV trajectory route design, RIS phase shift design, and active transmit beamforming problem to maximize the total system achievable rate under the given constraints. Thus, the reward for guided learning process should be consistent with the proposed multiple optimization objective. To achieve the objective of maximizing the total achievable rate, for constraints Eq. (16b)-(16e), we set a penalty that terminates the set of constraints if any of these constraints is not satisfied. Thus, the reward function is defined as

$$R^{(t)} = \begin{cases} -W, & \text{if } S_m = NS \\ R[t], & \text{otherwise,} \end{cases} \tag{22}$$

where NS denotes the negative state, that is, it does not satisfy any constraints in (15b)-(15g). W is a sufficiently large set of normal quantities to avoid not satisfying these constraints.

The agent does not know the state space transition probability matrix $P_{ss'}^a$ due to the uncertainty of UAV location and RIS phase shift configuration in this proposed MDP model. However, DRL algorithm is promising because it enables agents to control their actions without knowledge of the wireless environment.

B. MO-DDPG Algorithm Description

In this subsection, a multi-objective DDPG (MO-DDPG)-based optimization framework is investigated to tackle the jointly optimization of the UAV trajectory, transmission beamforming, and RIS phase shift configuration, which guaranteeing that all the system achievable rate and energy consumption are balanced. As seen in Fig. 2, we investigate the multi-objective DDPG (MO-DDPG) neural network to address this optimization issue. As we can be observed, the MO-DDPG neural network consists of two DNNs: the actor network and the critic network. The actor target network $\pi\left(\theta_a^{(target)}|s^{(t)}\right)$ and critic target network $Q\left(\theta_c^{(target)}|s^{(t)}, a^{(t)}\right)$ are both constructed using the dual network structure and the parameter update approach based on soft-strategy, which are used as the corresponding target values for the training network's parameter update, respectively. The actor network specifies the main policy for constructing the mapping from state to actions, and the critic network estimates the action values, where θ_a and θ_c correspond to the weighting and biasing parameters.

These actions are approximated using the actor network, so that the next non-convex optimization state is not required to find the maximum Q-value function. The following is an update strategy of the training critic network.

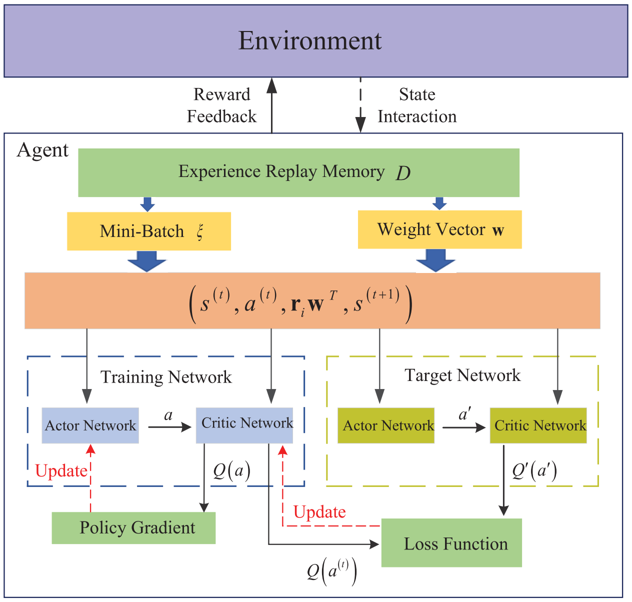


Fig. 2. Optimization Framework of MO-DDPG algorithm.

$$\theta_c^{(t+1)} = \theta_c^{(t)} - \mu_c \nabla_{\theta_c^{(train)}} \ell\left(\theta_c^{(train)}\right), \tag{23}$$

$$\ell\left(\theta_c^{(train)}\right) = \left(r^{(t)} + \gamma Q\left(\theta_c^{(target)} \mid s^{(t+1)}, a'\right) - Q\left(\theta_c^{(train)} \mid s^{(t)}, a^{(t)}\right)\right)^2, \tag{24}$$

where μ_c denotes the learning rate, a' is the action output and $\nabla_{\theta_c^{(train)}} \ell\left(\theta_c^{(train)}\right)$ denotes the update gradient in a specific time period. The weights of both the actor and critic networks are initialized by a truncated normal distribution centered at 0 with a normative deviation of $\sqrt{2/b_i}$, where b_i means the number of input cells in the weight tensor.

The training actor network is updated using the following policy gradient

$$\theta_a^{(t+1)} = \theta_a^{(t)} - \mu_a \nabla_a Q\left(\theta_c^{(target)} \mid s^{(t)}, a'\right) \times \nabla_{\theta_a^{(train)}} \pi\left(\theta_a^{(train)} \mid s^{(t)}\right), \tag{25}$$

where μ_a is the learning rate, $\nabla_a Q\left(\theta_c^{(target)} \mid s^{(t)}, a'\right)$ and $\nabla_{\theta_a^{(train)}} \pi\left(\theta_a^{(train)} \mid s^{(t)}\right)$ denote the gradient of target or target critic network with respect to their parameter $\theta_c^{(target)}$ and $\theta_c^{(train)}$, respectively. As can be seen from the above equation, the parameters update strategy is influenced by the gradien towards the action, thus ensuring that the next action is chosen in the direction preferred by the best action strategy, thus optimizing the Q-value function. Specifically, a target actor network $\theta_a^{(target)}$ and a target critic network $\theta_c^{(target)}$ are obtained by copying the parameters of these two training networks in the initialization phase.

For the multi-objective optimization problem, a novel multi-objective DRL-based framework is proposed, where instead of using all the information of the system as the input of the neural network, a small amount of information closely related to the decision is extracted to form the state vector. Since the value of an action relies on the preference between various optimization targets, a simple linear weighting approach is adopted to compute a weighted sum of the reward vector elements for a given weight $r = \mathbf{r}\mathbf{w}^T$, where $\mathbf{w} = \left[w_{R_K}, w_{E_k}\right]$ denotes the UAV energy constraint and the total rate optimization weights achievable by the system. The reward vector is then converted to scalar type to fit the network input form. With this design, the MO-DDPG algorithm is applicable to multi-objective optimization (MOO) problems with an arbitrary number of targets. In the considered algorithmic framework, all the involved weight coefficients in the interval [0, 1] is determined by the importance priority of the different optimization objectives.

The target value y_t for the target network is calculated as follows:

$$y_t = \mathbf{r}_t\mathbf{w}^T + \gamma \max_{a'} Q\left(\theta_a^{(target)}|s^{(t+1)}, a'\right). \tag{26}$$

The difference between the value of the objective function and the value of the Q-function given by the main critic network is also calculated during the optimization of the main critic network parameter setting. Then, gradient descent approach is adopted to train the main critic network, so as to minimize the loss function, i.e., the average square error of the diversity. To ensure space for sustained action is fully explored during the training process, exploration strategies are applied in the policy of actor transfer. In each decision step, the amount of network operations is chosen in a random process with expectation and variance $\varepsilon_i\sigma^2$, where ε_i is an adjustable parameter to attenuate the effect of the randomness of the actions in the training process. The complete pseudocode algorithm flow of the proposed multi-objective optimized DRL framework is given in Algorithm 1.

IV. Experimental Settings and Results

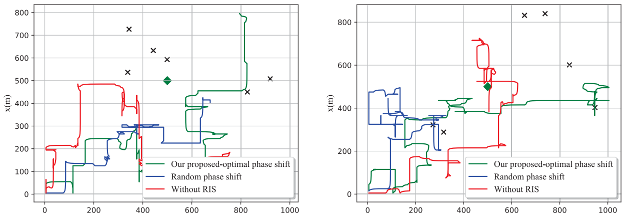
In this section, we provide experimental results under different performance indexes to quantify the performance of the proposed MO-DDPG algorithm for UAV trajectory design, satellite transmit beamforming design, and RIS passive phase shift design. In the simulation, the channel matrices $\mathbf{h}_k$ and $\mathbf{G}_r$ randomly generated following the above analytical channel model based on shadowed-Rician fading distribution [56].

In the initial phase of each algorithm execution, the UAV starts the communication task with 2D trajectory at random locations in the specified area. Table I lists the simulation system settings for the RIS-assisted IS-UAV-TNs. The envisaged UAV has a coverage radius of at least 10 km to serve the specific area.

In the MO-DDPG, both the proposed actor network and the critic network adopt the same fully connected neural network structure, which consists of an input layer for status information, an output layer to output the optimal action, two hidden layers and modular normalization components, like Fig. 3. We also train the policy network and the Q-network with AdamOptimizers, and randomly initializing the parameters of each DNNs according to the zero mean normal distribution.

Algorithm 1: MO-DDPG Algorithm

```
Initialize experience memory $D$, time slot count $T$,
  $\varepsilon = 0.99$, $\sigma^2 = 1.0$;
Initialize the train network and the target network,
  separately;
**Input**: $\mathbf{p}$, $\Phi$, $\mathbf{h}_k$, $\mathbf{G}_r$, $q_{S,n}$ and $q_{D,n}$;
**Output**: Optimal action $a_{opt}^{(t)}$, $Q$-value function;
**for** each episode **do**:
    Initialize state space as $s^{(0)} \in S$, $S \leftarrow s^{(0)}$;
    **for** $t = 0, 1, 2, \ldots T$ **do**:
        The central controller choose action $a^{(t)}$
          $= \pi\left(s^{(t)}|\theta_a^{(train)}\right) + \mathbb{R}$, where $\mathbb{R}$ is the random
          function for efficient exploration
          of the optimal action;
        Execute action at and limit UAV in designated area,
          observe reward $R^{(t)}$, and $s^{(t)}$ evolves into
          next state $s^{(t+1)}$;
        Save $(s^{(t)}, a^{(t)}, R^{(t)}, s^{(t+1)})$ into $D$;
        Randomly sample $\xi$ transitions form $D$;
        Process via DNN;
        Compute target value for the critic evaluation
          network by equation. (19);
        Update the parameters of the critic network by
          minimizing the critic loss
          $Loss\left(\theta_c^{(target)}\right) =$
          $\mathbb{E}\left[\left(y(t) - Q\left(\theta_c^{(train)}|s^{(t)}, a^{(t)}\right)\right)^2\right]$;
        Update the parameters of actor network with
          sampled policy gradients by
          $\nabla_{\theta_a^{(train)}} J =$
          $\frac{1}{\xi}\sum_{i=1}^{\xi} \nabla_a Q_\pi\left(\theta_c^{(train)}|\left(s^{(t)}, a^{(t)}\right)\right)\Big|_{a=\pi\left(s^{(t)}|\theta_a^{(train)}\right)}$
          $\nabla_{\theta_a^{(train)}}\pi\left(\theta_a^{(train)}|s^{(t)}\right)$;
        Soft-update the parameters of DDPG target
          networks by the following formula
          $\theta_c^{(target)} \leftarrow \tau_c\theta_c^{(train)} + (1-\tau_c)\,\theta_c^{(target)}$
          $\theta_a^{(target)} \leftarrow \tau_a\theta_a^{(train)} + (1-\tau_a)\,\theta_a^{(target)}$
        Update the state $s^{(t+1)}$;
    **end for**;
**end for**
```



(a) 2D Trajectory-Coverage area. (b) 2D Trajectory-Energy constraint.

Fig. 3. UAV trajectories design for different optimization purposes.

In addition, to demonstrate the effectiveness of the proposed framework, we examined whether to use the NOMA protocol and whether to deploy the performance of the system obtained

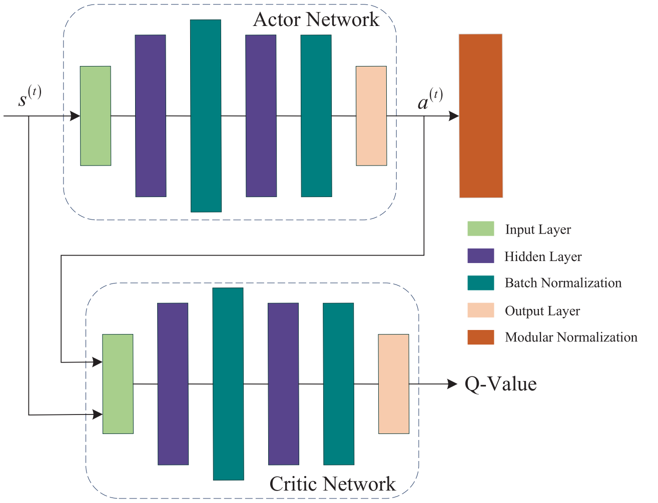


Fig. 4. Proposed DNN structure of the actor network and the critic network.

TABLE II

SIMULATION PARAMETERS SETTING

Simulation System Parameters	**Value**
Satellite type	GEO
Frequency band	$f = 2$ GHz
Maximal satellite gain	48 dB
Link bandwidth	$B = 15$ MHz
Noise power spectral density	-169 dBm/Hz
Variance of AWGN	1
Noise temperature	$300K$
3-dB angle	0.4°
UAV height	$\mathbf{H}_0 = 1000m$
UAV speed	$V_H = 10$m/s
Simulation DNN Parameters	**Value**
Number of time slots	100
Reward discount rate	0.99
Learning rate for target network	0.001
Learning rate for training network	0.001
Experience replay buffer size	100000
Number of episodes	10000
Optimizer	AdamOptimizer
Activation function	ReLu
Soft target update parameter	0.001
Amount of steps in per-episode	40000
Amount of experiences in the mini-batch	64

under the RIS scenario. Specifically, the following three benchmarks are considered:

• **Random phase shift**: At each time slot, the RIS phase shift values are randomly generated for each RIS sub-surface, which is called the UAV/R scheme.

• **Without RIS**: RIS is not deployed in the communication system, which is called UAV/NR scheme.

In our considered system model, we only assume UAV flying at a fixed altitude, so the Fig. 4 shows the 2D trajectory design for different optimization purposes. Compared to systems with RIS random phase shift and the UAV/NR scheme, RIS-assisted methods with optimized phase shift can usually cover a larger communication area within a limited energy constraint. Meanwhile, propulsion energy can be saved at the shortest distance. In the absence of RIS case, the UAV requires to seek suitable location to establish communication links to VUs, which obviously leads to low energy-efficient performance.

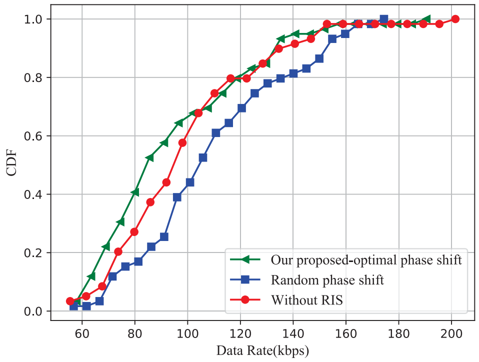


Fig. 5. The performance on data rate led by different solutions.

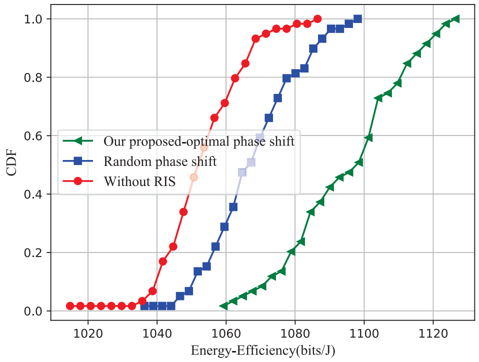


Fig. 6. The performance on energy-efficiency led by different solutions.

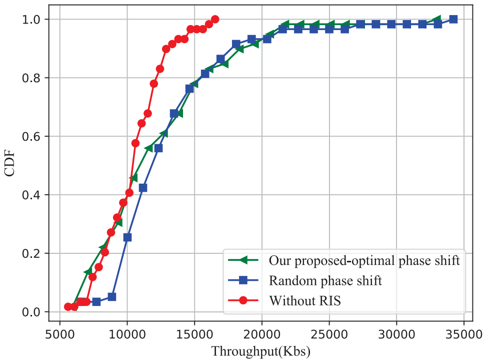


Fig. 7. The performance on throughput led by different solutions.

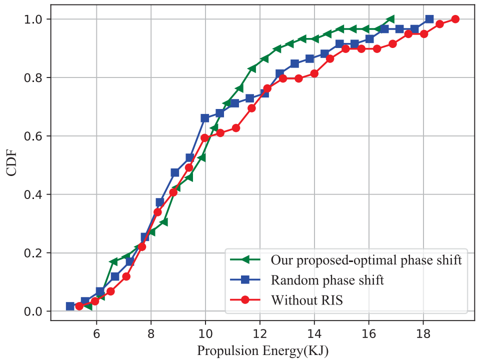


Fig. 8. The performance on propulsion energy led by different solutions.

In Fig. 5 - Fig. 8, the cumulative distribution functions (CDF) of data rate, energy efficiency, throughput, and propulsion efficiency is further depicted for three different communication systems. It can be seen that the RIS-assisted communication with optimized phase shift has a high performance in each test wireless scenario due to the random nature

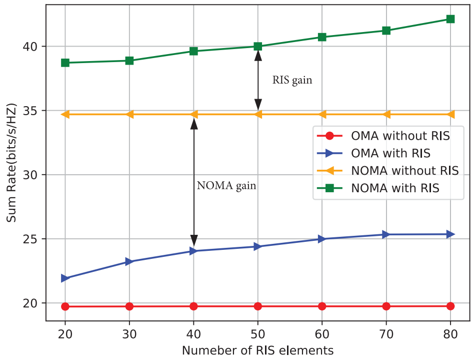


Fig. 9. Performance impact schematic for NOMA protocol and RIS deployment.

of the phase shift and the UAV location increases the difficulty of signal alignment, which increases the performance loss and reduces communication energy efficiency.

In addition, we examined the performance implemented under different scenarios using the NOMA protocol and deploying RIS. Specifically, the following three benchmarks below are taken into account.

• **Without RIS-NOMA**: In this scenario, RIS is not carried on the UAV. Only the direct link from satellite to VUs maintains communication and is not reflected by the RIS. Meanwhile, the user access scheme is through the NOMA protocol.

• **Without RIS-OMA**: Similar to the preceding communication scenario, except that the user access scheme is through the OMA protocol.

• **Optimal phase shift-OMA**: Optimized design of active and passive beamforming with MO-DDPG framework, but user access is OMA protocol.

It is worth our attention that the novel MO-DDPG algorithm framework proposed in this paper can also perform parameter optimization with the other three proposed benchmark schemes. Specifically, in the without RIS in NOMA-based scheme, adopting MO-DDPG algorithm to solve the sum-rate maximization problem by excluding the discrete phase shift of RIS into continuous action problems. For the random phase shift-OMA case, the optimization improvement issue can be settled by speculating that the satellite transmits with full power scheme or zero force transmitting scheme. With regard to the without RIS in OMA-based case, the problem can be settled by simply optimizing designing the UAV trajectory. According to Fig. 9, the sum rate grows with the number of RIS reflective elements, which is due to the higher beamforming gain caused by a larger meta-surface, the increased additional reflection links and the growth of spectral efficiency caused by the NOMA protocol applications.

V. Conclusion

In this paper, we proposed a novel RIS-assisted NOMA downlink IS-UAV-TNs system model to meet the sub-sequent practical applications requirements of the next generation wireless network construction. To maximize the system achievable data rate while minimizing the UAV energy consumption, a multi-objective DDPG-based optimization algorithm was developed to achieve online control of the UAV trajectory. By designing the reward function of the algorithm framework as a multi-dimensional vector corresponding to the multi-objective optimization, the powerful fitting capability of the neural network was exploited for the RIS phase shift and active transmit beamforming problems in the NOMA-based downlink system, while the UAV learned to find the joint optimization solution based on the weight parameters associated with the objective. It was worth noting that the multi-optimization objective framework could be extended to optimize problems with arbitrary number of objectives. The proposed MO-DDPG algorithm has achieved a balance between the speed of the training network and convergence to the local optimal solution. Simulation results show that in the added signal reflection mode through RIS, the system sum rate can be significantly improved. Moreover, it was shown that combining multiple UAVs through co-optimization trajectories has great advantages in improving the performance of complex communication tasks in IS-UAV-TNs.

References

[1] K. An et al., "Secure transmission in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3025–3037, Nov. 2016.

[2] Y. Su, Y. Liu, Y. Zhou, J. Yuan, H. Cao, and J. Shi, "Broadband LEO satellite communications: Architectures and key technologies," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 55–61, Apr. 2019.

[3] K. Guo et al., "Physical layer security for multiuser satellite communication systems with threshold-based scheduling scheme," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5129–5141, May 2020.

[4] L. Yang, Q. Zhu, X. Yan, S. Li, and H. Jiang, "Performance analysis of mixed PLC-FSO dual-hop communication systems," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 19307–19317, Oct. 2022.

[5] K. Guo, X. Li, M. Alazab, R. H. Jhaveri, and K. An, "Integrated satellite multiple two-way relay networks: Secrecy performance under multiple eves and vehicles with non-ideal hardware," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 2, pp. 1307–1318, Feb. 2023.

[6] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[7] S. Li, L. Yang, D. Benevides da Costa, and S. Yu, "Performance analysis of UAV-based mixed RF-UWOC transmission systems," *IEEE Trans. Commun.*, vol. 69, no. 8, pp. 5559–5572, Aug. 2021.

[8] X. Li et al., "Physical-layer authentication for ambient backscatter-aided NOMA symbiotic systems," *IEEE Trans. Commun.*, early access, Feb. 15, 2023, doi: 10.1109/TCOMM.2023.3245659.

[9] W. U. Khan, M. A. Javed, T. N. Nguyen, S. Khan, and B. M. Elhalawany, "Energy-efficient resource allocation for 6G backscatter-enabled NOMA IoV networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9775–9785, Jul. 2022.

[10] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[11] Z. Ding et al., "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[12] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual transceiver hardware impairments on cooperative NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 680–695, Jan. 2020.

[13] X. Li, M. Zhao, Y. Liu, L. Li, Z. Ding, and A. Nallanathan, "Secrecy analysis of ambient backscatter NOMA systems under I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12286–12290, Oct. 2020.

[14] L. Yang, F. Meng, M. O. Hasna, and E. Basar, "A novel RIS-assisted modulation scheme," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1359–1363, Jun. 2021.

[15] Y. Sun et al., "RIS-assisted robust hybrid beamforming against simultaneous jamming and eavesdropping attacks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9212–9231, Nov. 2022.

[16] L. Lv, Q. Wu, Z. Li, Z. Ding, N. Al-Dhahir, and J. Chen, "Covert communication in intelligent reflecting surface-assisted NOMA systems: Design, analysis, and optimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1735–1750, Mar. 2022.

[17] X. Yue, J. Xie, Y. Liu, Z. Han, R. Liu, and Z. Ding, "Simultaneously transmitting and reflecting reconfigurable intelligent surface assisted NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 189–204, Jan. 2023.

[18] H. Liu, X. Li, M. Fan, G. Wu, W. Pedrycz, and P. N. Suganthan, "An autonomous path planning method for unmanned aerial vehicle based on a tangent intersection and target guidance strategy," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3061–3073, Apr. 2022.

[19] C. Huang, G. C. Alexandropoulos, A. Zappone, C. Yuen, and M. Debbah, "Deep learning for UL&DL channel calibration in generic massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[20] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[21] M. T. Nguyen and L. B. Le, "Multi-UAV trajectory control, resource allocation, and NOMA user pairing for uplink energy minimization," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23728–23740, Dec. 2022.

[22] Q. Wang et al., "UAV-enabled non-orthogonal multiple access networks for ground-air-ground communications," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 3, pp. 1340–1354, Sep. 2022.

[23] N. Wang, F. Li, D. Chen, L. Liu, and Z. Bao, "NOMA-based energy-efficiency optimization for UAV enabled space-air-ground integrated relay networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4129–4141, Apr. 2022.

[24] Z. Na, Y. Liu, J. Shi, C. Liu, and Z. Gao, "UAV-supported clustered NOMA for 6G-enabled Internet of Things: Trajectory planning and resource allocation," *IEEE Internet Things J.*, vol. 8, no. 20, pp. 15041–15048, Oct. 2021.

[25] X. Pang, M. Sheng, N. Zhao, J. Tang, D. Niyato, and K.-K. Wong, "When UAV meets IRS: Expanding air-ground networks via passive reflection," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 164–170, Oct. 2021.

[26] C. Wang et al., "Covert communication assisted by UAV-IRS," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 357–369, Jan. 2023.

[27] L. Yang, Y. Yang, D. B. da Costa, and I. Trigui, "Outage probability and capacity scaling law of multiple RIS-aided networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 256–260, Feb. 2021.

[28] M. Asim, M. ELAffendi, and A. A. Abd El-Latif, "Multi-IRS and multi-UAV-assisted MEC system for 5G/6G networks: Efficient joint trajectory optimization and passive beamforming framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 4553–4564, Apr. 2023.

[29] Y. Liu et al., "An efficient power allocation algorithm for green reconfigurable intelligent surface assisted vehicular network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 23736–23749, Dec. 2022.

[30] F. Zhou, X. Li, M. Alazab, R. H. Jhaveri, and K. Guo, "Secrecy performance for RIS-based integrated satellite vehicle networks with a UAV relay and MRC eavesdropping," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 2, pp. 1676–1685, Feb. 2023.

[31] X. Li, Y. Zheng, M. Zeng, Y. Liu, and O. A. Dobre, "Enhancing secrecy performance for STAR-RIS NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2684–2688, Feb. 2023, doi: 10.1109/TVT.2022.3213334.

[32] X. Li, Z. Xie, Z. Chu, V. G. Menon, S. Mumtaz, and J. Zhang, "Exploiting benefits of IRS in wireless powered NOMA networks," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 175–186, Mar. 2022.

[33] X. Yue and Y. Liu, "Performance analysis of intelligent reflecting surface assisted NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2623–2636, Apr. 2022.

[34] H. Wang, C. Liu, Z. Shi, Y. Fu, and R. Song, "Power minimization for two-cell IRS-aided NOMA systems with joint detection," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1635–1639, May 2021.

[35] L. Lv, Q. Wu, Z. Li, Z. Ding, N. Al-Dhahir, and J. Chen, "Achieving covert communication by IRS-NOMA," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Jul. 2021, pp. 421–426.

[36] Y.-L. Lan, F. Liu, W. W. Y. Ng, M. Gui, and C. Lai, "Multi-objective two-echelon city dispatching problem with mobile satellites and crowd-shipping," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15340–15353, Sep. 2022.

[37] J. Watts, F. Van Wyk, S. Rezaei, Y. Wang, N. Masoud, and A. Khojandi, "A dynamic deep reinforcement learning-Bayesian framework for anomaly detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 22884–22894, Dec. 2022.

[38] M. Luo et al., "Fleet rebalancing for expanding shared e-Mobility systems: A multi-agent deep reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 3868–3881, Apr. 2023.

[39] C. Zhong et al., "Deep reinforcement learning-based optimization for IRS-assisted cognitive radio systems," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3849–3864, Jun. 2022.

[40] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042–2055, Jul. 2021.

[41] G. Lee, M. Jung, A. T. Z. Kasgari, W. Saad, and M. Bennis, "Deep reinforcement learning for energy-efficient networking with reconfigurable intelligent surfaces," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[42] W. Xu, J. Zhang, P. Zhang, and C. Tellambura, "Outage probability of decode-and-forward cognitive relay in presence of primary user's interference," *IEEE Commun. Lett.*, vol. 16, no. 8, pp. 1252–1255, Aug. 2012.

[43] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Aerial intelligent reflecting surface: Joint placement and passive beamforming design with 3D beam flattening," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4128–4143, Jul. 2021.

[44] X. Yue et al., "Outage behaviors of NOMA-based satellite network over shadowed-Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6818–6821, Jun. 2020.

[45] M. K. Arti, "Channel estimation and detection in satellite communication systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10173–10179, Dec. 2016.

[46] K. Guo, M. Lin, B. Zhang, W.-P. Zhu, J.-B. Wang, and T. A. Tsiftsis, "On the performance of LMS communication with hardware impairments and interference," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1490–1505, Feb. 2019.

[47] G. Zheng, P. D. Arapoglou, and B. Ottersten, "Physical layer security in multibeam satellite systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 852–863, Feb. 2012.

[48] Z. Lin, M. Lin, W.-P. Zhu, J.-B. Wang, and J. Cheng, "Robust secure beamforming for wireless powered cognitive satellite-terrestrial networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 2, pp. 567–580, Jun. 2021.

[49] K. Guo, "Performance analysis of hybrid satellite-terrestrial cooperative networks with relay selection," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9053–9067, Aug. 2020.

[50] H. Liu et al., "Effective capacity analysis of STAR-RIS assisted NOMA networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1930–1934, Sep. 2022.

[51] L. Lv, H. Jiang, Z. Ding, Q. Ye, N. Al-Dhahir, and J. Chen, "Secure non-orthogonal multiple access: An interference engineering perspective," *IEEE Netw.*, vol. 35, no. 4, pp. 278–285, Jul. 2021.

[52] L. Lv, H. Jiang, Z. Ding, L. Yang, and J. Chen, "Secrecy-enhancing design for cooperative downlink and uplink NOMA with an untrusted relay," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1698–1715, Mar. 2020.

[53] K. An, M. Lin, W.-P. Zhu, Y. Huang, and G. Zheng, "Outage performance of cognitive hybrid satellite–terrestrial networks with interference constraint," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9397–9404, Nov. 2016.

[54] Z. Lin, M. Lin, J.-B. Wang, T. de Cola, and J. Wang, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 3, pp. 657–670, Jun. 2019.

[55] H. Mei, K. Yang, Q. Liu, and K. Wang, "3D-trajectory and phase-shift design for RIS-assisted UAV systems using deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 3020–3029, Mar. 2022.

[56] N. I. Miridakis, D. D. Vergados, and A. Michalas, "Dual-hop communication over a satellite relay and shadowed Rician channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4031–4040, Sep. 2015.

Kefeng Guo received the B.S. degree from the Beijing Institute of Technology, Beijing, China, in 2012, and the Ph.D. degree from Army Engineering University, Nanjing, China, in 2018.

He is currently a Lecturer with the School of Space Information, Space Engineering University. He is also an Associate Professor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. He has authored or coauthored nearly 70 research papers in international journals and conferences. His research interests include cooperative relay networks, MIMO communications systems, multiuser communication systems, satellite communication, hardware impairments, cognitive radio, NOMA technology, and physical layer security. He was a recipient of Exemplary Reviewer for IEEE TRANSACTIONS ON COMMUNICATIONS in 2022. He was a recipient of the Outstanding Ph.D. Thesis Award of Chinese Institute of Command and Control in 2020. He was also a recipient of the Excellent Ph.D. Thesis Award of Jiangsu Province, China, in 2020. He has been a TPC Member of many IEEE sponsored conferences, such as IEEE ICC, IEEE GLOBECOM, and IEEE WCNC. He also serves as an Editor on the Editorial Board for the *EURASIP Journal on Wireless Communications and Networking*. He was the Guest Editor for the Special Issue on Integration of Satellite-Aerial-Terrestrial Networks of *Sensors* and also the Guest Editor for the Special Issue on Recent Advances and Challenges of Satellite and Aerial Communication Networks of *Electronics*.

Min Wu received the M.S. degree from Space Engineering University, Beijing, China, in 2021, where she is currently pursuing the Ph.D. degree. Her research interests include satellite-terrestrial networks, RIS-assisted wireless communication systems, multiuser communication systems, and deep reinforcement learning.

Xingwang Li (Senior Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China in 2010 and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2015.

From 2010 to 2012, he was with Comba Telecom Ltd., Guangzhou, China, as an Engineer. He spent one year from 2017 to 2018 as a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He is currently an Associate Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include wireless communication, intelligent transport systems, artificial intelligence, and the Internet of Things. He has served as many TPC members, such as the IEEE Globecom, IEEE ICC, IEEE WCNC, IEEE VTC, and IEEE ICCC. He has also served as the Co-Chair for the IEEE/IET CSNDSP 2020 of the Green Communications and Networks Track. He also serves as an Editor on the Editorial Board for IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE SYSTEMS JOURNAL, PHYSICAL COMMUNICATION. He was the Guest Editor for the Special Issue on Computational Intelligence and Advanced Learning for Next-Generation Industrial IoT of IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, "Recent Advances in Physical Layer Technologies for 5G-Enabled Internet of Things" of the Wireless Communications and Mobile Computing. He is also the Co-Chair of IEEE/IET CSNDSP 2020 of the Green Communications and Networks Track.

Houbing Song (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in August 2012.

He is currently a Tenured Associate Professor, the Director of the NSF Center for Aviation Big Data Analytics (Planning), the Associate Director for Leadership of the DOT Transportation Cybersecurity Center for Advanced Research and Education (Tier 1 Center), and the Director of the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us), University of Maryland Baltimore County (UMBC), Baltimore, MD. Prior to joining UMBC, he was a Tenured Associate Professor in electrical engineering and computer science with Embry-Riddle Aeronautical University, Daytona Beach, FL. He is an editor of eight books, the author of more than 100 articles, and an inventor of two patents. His research interests include cyber-physical systems/the Internet of Things, cybersecurity and privacy, and AI/machine learning/big data analytics. His research has been sponsored by federal agencies (including National Science Foundation, National Aeronautics and Space Administration, U.S. Department of Transportation, and Federal Aviation Administration, among others) and industry. His research has been featured by popular news media outlets, including IEEE GlobalSpec's Engineering360, Association for Uncrewed Vehicle Systems International (AUVSI), Security Magazine, CXOTech Magazine, Fox News, U.S. News & World Report, The Washington Times, and New Atlas. He is an IEEE Fellow (for contributions to big data analytics and integration of AI with Internet of Things) and an ACM Distinguished Member (for outstanding scientific contributions to computing). He has been an ACM Distinguished Speaker since 2020 and an IEEE Vehicular Technology Society (VTS) Distinguished Lecturer since 2023. He has been a Highly Cited Researcher identified by Clarivate^{TM} in 2021 and 2022 and a Top 1000 Computer Scientist identified by Research.com. He received Research.com Rising Star of Science Award in 2022 (World Ranking: 82; U.S. Ranking: 16). He was a recipient of more than ten best paper awards from major international conferences, including IEEE CPSCom-2019, IEEE ICII 2019, IEEE/AIAA ICNS 2019, IEEE CBDCom 2020, WASA 2020, AIAA IEEE DASC 2021, IEEE GLOBECOM 2021, and IEEE INFOCOM 2022. He has been serving as an Associate Editor for IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS since 2020, IEEE INTERNET OF THINGS JOURNAL since 2020, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS since 2021, and IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE since 2023. He was an Associate Technical Editor of *IEEE Communications Magazine* (2017–2020).

Neeraj Kumar (Senior Member, IEEE) is currently a Full Professor and the Dean of DCT with the Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology (Deemed to be University), Patiala, India. He is also an Adjunct Professor with King Abdul Aziz University, Jeddah, Saudi Arabia, and Newcatle University, U.K. He has published more than 500 technical research papers in top-cited journals and conferences, which are cited more than 39000 times from well-known researchers across the globe with current H-index of 110. He has guided many research scholars leading to M.E./M.Tech. and Ph.D. His research is supported by funding from various competitive agencies across the globe. His broad research interests include green computing and network management, the IoT, big data analytics, deep learning, and cyber-security. He has also edited/authored ten books with international/national publishers, such as IET, Springer, Elsevier, and CRC. He has secured research funding of around 1 Million Euro from Government of India, and industries in the area of smart grid, blockchain, cyber-security, and network management. He is a Highly-Cited Researcher from WoS in 2019, 2020, and 2021. He is a consultant in various industry and government sponsored projects in China, Saudi Arabia, and Europe. He has executed various international projects in Austria, Poland, Saudi Arabia, Europe, and China. He has supervised more than 15 Ph.D. and 25 M.E./M.Tech. students. He is serving as an Editor for *ACM Computing Survey*, IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, *Computer Communication* (Elsevier), and international journal (Wiley).