Open Access

Energy efficiency performance in RIS-based integrated satellite-aerial-terrestrial relay networks with deep reinforcement learning



Jiao Li¹, Huajian Xue^{2*}, Min Wu^{3*}, Fucheng Wang², Tieliang Gao¹ and Feng Zhou⁴

*Correspondence: xuehj@tlu.edu.cn; minwu2024@163.com; 1800022837@pku.edu.cn

 Key Laboratory of Data Analysis and Financial Risk
 Prediction, Xinxiang University, Xinxiang 453000, China
 ² Tongling University, Tongling 244000, China
 ³ School of Space Information, Space Engineering University, Beijing 101400, China
 ⁴ School of Information Technology, Yancheng Institute of Technology, Yancheng 224000, China

Abstract

Integrated satellite–aerial–terrestrial relay networks (ISATRNs) play a vital role in nextgen networks, particularly those with high-altitude platforms (HAP). This study introduces a new model for hybrid optical/RF-based HAP-enabled ISATRNs, incorporating reconfigurable intelligent surfaces (RIS) on unmanned aerial vehicles (UAVs) to optimize access in dense urban areas. Non-orthogonal multiple access is employed for improved spectrum efficiency. The objective is to jointly optimize UAV trajectory, RIS phase shift, and active transmit beamforming while considering energy consumption. A deep reinforcement learning approach using LSTM-DDQN framework is proposed. Numerical results show the effectiveness of our algorithm over traditional DDQN, with higher single-step exploration reward and evaluation metrics.

Keywords: Satellite–aerial–ground integrated networks (SAGINs), Hybrid FSO/RF links, Reconfigurable intelligent surface (RIS), Long short-term memory-deep deterministic policy gradient (LSTM-DDPG), Deep reinforcement learning (DRL)

1 Introduction

With the development of the tourism industry and aerial communication, aerial technology has brought great innovation to the tourism industry, which provides unique and unprecedented perspectives and experiences for tourists, and also brings new business opportunities to the tourism industry [1]. However, the existing network cannot effectively support the development of aerial technology, on this foundation, integrated satellite–aerial–terrestrial relay networks (ISATRNs) have gained significant attention as a potential infrastructure from academia and industry to meet the increasing demands for capacity and reliability, which can satisfy the requirements of the urban lives and tourism industry [2, 3]. There networks utilize high-altitude-platforms (HAPs) and unmanned aerial vehicles (UAVs) to expand a wide service coverage area in various broadband wireless communication applications [4–6]. HAPs, including airships and aircraft, operate at altitudes of 20–25 km in the stratosphere, providing increased maneuverability. In contrast, unmanned aerial vehicles (UAVs) or drones operate at altitudes ranging from a few tens to approximately 100 ms above the ground. UAVs offer a favorable air-to-ground



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

channel for direct line-of-sight (LoS) communication with ground cellular equipment. This enables reliable wireless connectivity, especially during emergency situations when terrestrial networks are overloaded, incapacitated, or completely destroyed. By utilizing signal amplification relays, UAVs can swiftly establish communication links, ensuring seamless connectivity. The deployment of HAPs and UAVs plays a crucial role in maintaining effective wireless communication. These aerial platforms effectively bridge communication gaps, facilitating uninterrupted connectivity in challenging environments. This capability greatly contributes to efficient emergency response and enhances overall communication systems.

On the one hand, free-space optical (FSO) communication is expected to fulfill the requirements for next-generation ISATRNs that demand high-speed and highly secure connections. FSO technology offers several advantages, such as high-bandwidth capacity, no spectrum license requirements, high immunity to interference, compatibility with radio frequency (RF) communication, and easy installation. However, FSO transmission is vulnerable to atmospheric turbulence, which can significantly reduce system performance. Additionally, because FSO communication depends on free space for signal transmission, obstacles can obstruct the optical signals, leading to non-line-of-sight transmission challenges [7]. To overcome the connectivity challenges in the "last-mile," introducing an asymmetric hybrid multi-hop RF/FSO transmission infrastructure is being considered as a promising solution within existing wireless systems. By leveraging the strengths of both RF and FSO technologies, this system can provide extensive coverage and high-bandwidth connectivity, contributing to the construction of a robust wireless network while mitigating the limitations of FSO communications. The hybrid infrastructure can enable seamless switching between RF and FSO technologies, maximizing the benefits of each and reducing dependence on a single type of communication technology. In [8], the authors discussed the effect of different system parameters on the system outage probability and average bit error rate in UAV-enabled multi-hop FSO-based transmission systems

Considering that the downlink of ISATRNs can be severely degraded in urban environments because the channel suffers from severe penetration loss. Meanwhile, to facilitate the implementation of emerging services and industrial advancement, future wireless communication networks will offer converged services encompassing communication, perception, and computation, commonly referred to as communicationsensing-computing. As a significant technology option for 6 G, reconfigurable intelligent surfaces (RISs) demonstrate the capability in communication, perception, and computation domains. Moreover, they hold the potential to establish an integrated system that combines communication, sensing, and computation, which has garnered substantial attention in recent years. RIS, as an extension of metasurface materials, has been proposed to be applied in the field of communication, whose overall purpose is to intelligently reconfigure the signaling wireless environment [9-11]. Besides, non-orthogonal multiple access (NOMA) technology makes efficient use of the available spectrum by using non-orthogonal transmission with allocated user transmit power at the transmitter side and applying successive interference cancelation (SIC) at the receiver side [12, 13]. The authors in [14] conducted the performance of downlink MIMO-NOMA systems with discrete phase-shifted distributed RIS support. A comprehensive framework was developed to facilitate three access modes, aiming to mitigate intra-cluster interference effectively. Theoretical analysis was explored to derive the outage probability of the involved cascade channel under various scenarios. Additionally, experimental simulations were performed to validate the efficacy of the proposed scheme. Inspired by the strategy of fixed multiple RISs working together, researchers have gradually shifted their attention to mobile RISs for higher degrees of freedom. A notable approach, as discussed in [15], entails the deployment of RIS on mobile UAVs. The authors provided empirical evidence to support the efficacy of this approach, exemplifying the considerable advantages offered by utilizing UAVs equipped with RIS to enhance signal reflection coverage. According to the experimental results, the mobile RIS scheme can provide more flexible reflection strategies and the application scenarios can be more generalized. In mobile multiplexed schemes that combine RIS with FSO/RF hybrid transmission techniques, the integration of NOMA within the ISATRN necessitates meticulous attention to several key factors during the signaling process [16]. These include the design of active beamforming for transmission, RIS phase-shift control, and UAV trajectory. Taking these factors into account is crucial to achieve optimal system performance in terms of signal transmission and coverage, but due to their high-dimensionality and non-convexity, they are difficult to solve using traditional iterative optimization algorithms or require significant computational resources. Simultaneously, deep reinforcement learning (DRL) techniques are demonstrating remarkable capabilities in various domains, owing to their real-time interaction with the environment. As a result, researchers are progressively shifting their focus toward leveraging DRL for the optimization of communication performance. The authors in [17] have addressed the problem of performance optimization in a scenario where UAVs move within RIS-assisted ISATRNs. They have formulated a multi-objective optimization problem and designed a multidimensional reward function to optimize the phase configuration objective. The experimental results indicate that this approach can be readily extended to other scenarios. To the best of our knowledge, there are no existing studies on accurate signal transmission from satellites to ground devices using RIS-aided ISATRNs incorporating hybrid FSO/RF modes. Furthermore, joint beamforming design using an extended DRL algorithmic framework has not been explored. These challenges motivated the focus of our research paper and contributed significantly to the advancement of this field.

In the context of mobile multiple RISs and FSO/RF hybrid scenarios, optimizing the signal transmission of ISATRNs using non-orthogonal multiple access (NOMA) requires careful consideration of several factors [16]. These include the design of active beamforming for transmission, RIS phase-shift control, and UAV trajectory planning. Taking these factors into account is crucial to achieve optimal system performance in terms of signal transmission and coverage. Deep reinforcement learning (DRL) that enables the discovery of optimal policies by learning from environmental interactions through a trial-and-error process. Instead, it learns from rewards and penalties received during its interactions with the environment. The authors in [17] achieved multi-objective optimization problems by designing multidimensional vectors that corresponded to the reward function. These vectors aimed to optimize various factors, including active beamforming, passive beamforming, and UAV trajectory constraints. The authors in [18] mentioned the use of a dual deep neural network architecture in a supervised learning

situation to optimize the RIS phase-shift matrix, and simulations verified the effectiveness. In [19], the authors discussed the RIS phase-shift optimization problems by using the two-delay depth deterministic (TD3)-based method. Numerical results showed that the transmit power of this algorithm is essentially the same as the lower bound of the transmit power of the streaming optimization algorithm and significantly reduced the computational delay. In [20], the authors investigated the use of DRL algorithms to solve RIS-assisted multiuser full-duplex secure communication systems with the objective of maximizing the total secrecy rate. Through the existing literature, we can summarize that DRL can obtain satisfactory solutions through iterative interaction with and learning from dynamic environments. In addition, an appropriate neural network architecture can greatly improve the performance of DRL-based methods and speed up the convergence of neural networks.

This paper focuses on the modeling of downlink transmission in ISTRNs using a hybrid FSO/RF transmission approach. To improve the line-of-sight link and enhance the service quality for ground users, we assume that the RIS is installed on the UAV's side to provide accurate reflection services. The proposed system model involves joint optimization of the transmit beamforming and the RIS phase-shift matrix. We propose a modified DRL-based LSTM-DDQN algorithm to solve this optimization problem. Specifically, the main focus of this paper is as follows.

- Firstly, we present a system that enables high-capacity wireless communication from a satellite to the ground by employing a hybrid FSO/RF-based transmission mode. To address the challenges of operating in an ultra-dense environment, we deploy a RIS array on a UAV. This RIS array reflects the signal from HAPs back to the ground equipment, effectively mitigating the adverse effects of the high-density environment. Additionally, the HAP utilizes NOMA technology for transmission to the ground, further enhancing the system's efficiency.
- Secondly, to tackle the challenges related to joint active beamforming, and passive beamforming in RIS-assisted ISATRNs, we propose the utilization of an advanced DRL framework that surpasses the limitations of conventional approaches. Specifically, we introduce an enhanced LSTM-DDQN algorithm that enables the optimization of problems in both discrete and continuous environments. This algorithm provides a comprehensive solution for the joint optimization, allowing for efficient system performance.
- Finally, we conduct a thorough validation analysis to showcase the superiority of our proposed optimization algorithm in the given system model. Through extensive numerical experiments, we provide compelling evidence that the DRL-based solution outperforms various benchmark solutions. Additionally, we perform a comparative analysis between the DDQN algorithm and the LSTM-DDQN algorithm to evaluate their respective performances. Specifically, we observe a significant 11% improvement in the reward value when utilizing the LSTM-DDQN algorithm compared to the traditional DDQN algorithm.

The subsequent sections of this paper are organized as follows. In Sect. 2, we offer a comprehensive overview of the NOMA-based system model under consideration,

including a detailed description and formulation of the optimization problem. In Sect. 3, our focus shifts toward addressing the energy efficiency problem using the DRL-based algorithm. Next, in Sect. 4, we present simulation results that highlight the advantages of the extended LSTM-DDQN framework. Finally, in Sect. 5, we provide concluding remarks summarizing the key findings and potential future research directions. Please refer to Table 1 for additional abbreviations used in this paper. It provides a comprehensive list of abbreviations and their corresponding meanings.

2 Modeling of proposed system and optimization problem

As shown in Fig. 1, a new downlink mixed FSO/RF-enabled ISATRNs system is presented for reliable services from satellites to ground equipments (UEs). This mixed system uses FSO technology for satellite-to-HAP transmission and RF for HAP-to-UEs communication. The HAP utilizes NOMA-enabled beamforming with a uniform linear array (ULA) composed of N_a apertures and N_h antennas to collect optical signals from the satellite. Conversely, the UAV is equipped with a single RIS and serves $K(K \le N_h)$ mobile UEs [21]. Based on the three-dimensional (3D) Cartesian coordinate system, the HAP position is fixed at $\mathbf{L}_H = [x_H, y_H, z_H]^T$, while the position of each target UEs is denoted by $\mathbf{b}_k = [x_k, y_k]$.

	Catallita communication
	Satellite continunication
FSO	Free-space optics
RF	Radio frequency
STNs	Satellite terrestrial networks
RIS	Reconfigurable intelligent surfaces
SDMA	Space division multiple access
NOMA	Non-orthogonal multiple access
OP	Outage probability
IoT	Internet of things
ML	Machine learning
DRL	Deep reinforcement learning
DDQN	Double deep Q-network
LSTM	Long short-term memory
DDPG	Deep deterministic policy gradient
MIMO	Multiple input multiple output
SER	System ergodic rate
ULA	Uniform linear array
WDM	Wavelength division multiplexing
EGC	Equal-gain combing
AWGN	Additive white Gaussian noise
BF	Beamforming
DF	Decode-and-forward
SR	Shadowed-Rician
SNR	Signal-to-noise ratio
FSL	Free-space loss
GEO	Geosynchronous earth orbit
i.d	Independent and identically distribution

Table 1 Abbreviations



Fig. 1 Proposal of hybrid FSO/RF with NOMA-based RIS-UAV relay for HAP-based ISATRNs

The fixed-wing UAV needs to move from its starting point to its destination within a total time of T_t , divided into N time slots of duration $\varsigma_t = T_t/N$. Under the assumptions of block fading and constant channel conditions within each time slot, the 3D spatial position of the UAV can be represented as $\mathbf{q}[n] = [x[n], y[n], z[n]]^T$, $n \in N = \{0, ..., N\}$. Additionally, our proposed system assumes perfect hardware impairments and takes into account other related errors to simplify the analysis.

2.1 Signal transmission model

The transmission process in the proposed system can be analyzed as two separate modes. Firstly, during the first phase, the satellite uses optical communication terminals and telescopes to transmit optical signals to the HAP. The HAP then receives these signals through its optical receiving apertures and converts them into electrical RF signals. This process is essential for enabling the use of RF communication in the next transmission phase, which involves the HAP using the ULA and NOMA-enabled beamforming techniques to transmit RF signals to the UEs on the ground. Thus, the converted RF signal in the n_{α} th aperture of the HAP can be denoted as

$$y_{HAP,q} = \sqrt{P_S \eta_{oe} h_{SH} x_S + N_{n_a}}, n_a = 1, 2, ..., N_A,$$
(1)

where P_S denotes the satellite launch power, η_{oe} is the conversion coefficient, x_S represents the intensity-modulated optical signal emitted by the satellite, and n_q can be regarded as additive white Gaussian noise (AWGN) satisfying $N_{n_a} \sim C\mathcal{N}(0, \sigma_a^2)$. In addition, h_{SH} is the scalar channel fading coefficient from the satellite to the HAP,

which models atmospheric turbulence and pointing errors in the FSO channel. Based on these parameters, the output electrical SNR of the combined signal at the HAP can be expressed as

$$\gamma_H = \frac{P_S \eta_{oe}^2 h_{EGC}^2}{N_a n_a} \triangleq \bar{\gamma}_H h_{EGC}^2, \tag{2}$$

where $h_{EGC} = \sum_{n_a=1}^{N_a} h_{SH}$ represents the scalar channel fading coefficient of the receive aperture ensemble and $\bar{\gamma}_H = \left[\left(P_S \eta_{oe}^2 \right) / N_a n_a \right]$ is the average SNR. The SER for the FSO link can be expressed as

$$R_F = B_F E \left[\log_2(1 + \gamma_H) \right] = B_F E \left[\log_2 \left(1 + \bar{\gamma}_H h_{\text{EGC}}^2 \right) \right], \tag{3}$$

where B_F denotes the FSO transmission link bandwidth.

In the second phase, the HAP uses the NOMA-enabled transmission technique to transmit RF signal $x_k(t)$ with $E[|x_k(t)|^2] = 1$ to the *k*th UEs at the *t*th time slot [22]. Recognizing that most users are in remote areas, we utilize a UAV carrying a RIS as a movable reflective platform for service provision instead of traditional ground-fixed base stations [23]. We define the phase-shift matrix applied at the RIS as $\Phi \triangleq [\beta_1 e^{i\theta_1}, \beta_2 e^{j\theta_2}, ..., \beta_m e^{i\theta_m}, ..., \beta_M e^{i\theta_M}]^T$, where $\beta_m \in \mathcal{A}$ and $\theta_m \in \Theta$ represent the reflection amplitude and phase shift, respectively, of the *m*th element among the *M* total reflective elements [24, 25]. Let $\mathbf{h}_{\text{HR}} \in \mathbb{C}^{M \times N_H}$ and $\mathbf{g}_{\text{RG}} \in \mathbb{C}^{1 \times M}$ be defined as the channel gains from the HAP to the RIS and from the RIS to the *k*th IoT device, respectively. The gain of the direct transmission link from the HAP to the *k*th IoT device can be expressed as $\mathbf{G}_k^H \in \mathbb{C}^{N_H \times 1}$. Then, let all involved channel gains, including the direct channel and the reflection channel, denoted as $\mathbf{h}_1^H, ..., \mathbf{h}_K^H$, where the combined channel coefficient through the RIS experienced by the *k*th UEs can be expressed as $\mathbf{h}_k^H = \mathbf{h}_{\text{HR}}^H \Phi \mathbf{g}_{\text{RG}} + \mathbf{G}_k^H$. Thus, the received signal can be formulated as

$$y_k = \left(\sum_{k}^{K} \mathbf{w}_k x_k\right) \mathbf{h}_k^H + n_k, \forall k,$$
(4)

where $\mathbf{w}_k \in \mathbb{C}^{N_H \times 1}$ denotes the transmit beamforming vector at the HAP and n_k represents AWGN satisfying $n_k \sim C\mathcal{N}(0, \sigma_k^2)$. To maintain the transmit power at the HAP, the following constraint must also be imposed:

$$E\left\{tr\left\{\mathbf{wx}(\mathbf{wx})^{H}\right\}\right\} \le P_{t},\tag{5}$$

where P_t represents the total transmit power. The signal-to-interference-plus-noise ratio (SINR) $\tilde{\gamma}_{k,i}$ of the signal for the *k*th UEs decoded at the *i*th UEs ($i = 1, ..., k, ..., K, k \le i$) can be expressed as

$$\tilde{\gamma}_{k,i} = \frac{\left|\mathbf{h}_{i}^{H}\mathbf{w}_{k}\right|^{2}}{\sum_{j=k+1}^{K}\left|\mathbf{h}_{i}^{H}\mathbf{w}_{j}\right|^{2} + \sigma_{k}^{2}}.$$
(6)

Without loss of optimality, the condition for being able to perform SIC in this model is that the *i*th UE is able to successfully decode the signal for the *k*th IoT device ($k \le i \le K$);

then, the *k*th IoT device's SINR $\tilde{\gamma}_k$ can be represented as $\tilde{\gamma}_k = \min \{\tilde{\gamma}_{k,i}\}$. In other words, when $\tilde{\gamma}_k$ is the minimal value in the $\tilde{\gamma}_{k,i}$ set, it is guaranteed that the signal for the *k*th UE can be decoded by the user at the *i*th UE no matter what value it takes to ensure successful implementation of SIC. Thus, the SER for the RF transmission link can be obtained as

$$R_{R} = B_{R} \sum_{k=1}^{K} R_{k} = B_{R} \sum_{k=1}^{K} E\left[\log_{2}(1+\tilde{\gamma}_{k})\right],$$
(7)

where B_R denotes the RF-based transmission mode bandwidth. In this setup, the HAP employs the decode-and-forward (DF) protocol. This protocol involves the HAP buffering the received signal and subsequently forwarding it after amplification. Thus, the SER can be defined as

$$R(t) = C(\mathbf{q}, \mathbf{w}, \boldsymbol{\Phi}, \mathbf{h}_k, h_{\text{SH}}) = \min(R_F, R_R).$$
(8)

2.2 Channel model

1) FSO-based satellite-to-HAP channel model: FSO communication technology is a potential solution for downlink transmission from the satellite to the HAP, as it provides ultrahigh-capacity and secure transmission capabilities over long distances. In contrast to traditional optical receivers that rely on single-aperture solutions, the HAP considered in this paper is equipped with multiple receive apertures, denoted as N_a . The N_a FSO channels can be modeled using various parameters, including the channel attenuation coefficient, pointing error, and link distance, which can be expressed as

$$h_{\rm SH} = h_l h_a,\tag{9}$$

where $h_l = \frac{1}{2}(G_T + G_R - A_{\rm FS} - A_{\rm ATM} - L_{\rm loss} - M_S)$, with G_T , G_R , $A_{\rm FS}$, $A_{\rm ATM}$, $L_{\rm loss}$, and M_S being the transmit antenna gain, receive antenna gain, free-space loss, atmospheric attenuation, lenses loss, and system margin, respectively. The Gamma–Gamma distribution is commonly used to represent the fading parameter h_a in an FSO link affected by atmospheric turbulence. It is specifically designed to capture the impact of turbulence on optical wave propagation. This statistical model is preferred over other turbulence models because it incorporates both large-scale fading coefficients and small-scale fading factors, which are correlated with atmospheric channel characteristics. Unlike alternative models such as log-normal and exponential distributions [26], the Gamma–Gamma distribution offers several advantages. It accurately describes the intensity of atmospheric turbulence and predicts fluctuations in optical intensity under various turbulence conditions, including both strong and weak turbulence. This makes it a suitable choice for modeling the atmospheric channel. Additionally, by estimating the parameters of the Gamma–Gamma distribution, we can obtain valuable information about the atmospheric channel state eaily.

2) *RF-based RIS-assisted HAP-to-IoT device channel model*: To optimize the performance of multiuser communication systems that employ RIS, a wide range of innovative communication technologies and techniques must be integrated. By incorporating advanced approaches, we can maximize the system's overall performance and fully leverage its capabilities. Some of these approaches include beamforming design, optimal resource allocation, and user grouping scheduling, all of which are reliant on accurate channel state information (CSI). However, the presence of reflective elements in the RIS introduces complexities in channel estimation, as simultaneous estimation is required for several channels, including the direct link between the HAP and each ground UE, the HAP-to-RIS channel, and the RIS-to-UEs channel. Accurate channel estimation in multiuser scenarios is challenging but crucial. With a suitable channel estimation technique, we can assume that the global CSI is known for reception purposes, making it easier to implement effective optimization strategies for resource allocation, and user scheduling. Ultimately, these strategies can enhance the performance of RIS-assisted multiuser communication systems. The HAP-to-UEs direct channel vector follows the Nakagami-*m* distribution and can be expressed as

$$\mathbf{G}_{H,k} = C_{H,k} g_{H,k} \mathbf{A}(\varphi_h, \theta_h), \tag{10}$$

where $g_{H,k}$ is the random variable with channel fading severity level. $C_{H,k}$ represents the loss component between the HAP and the UEs, which can be computed as

$$C_{H,k} = G_H + R_k + \frac{1}{2} (20 \lg \lambda_F - 10\varpi \lg D_{H,k} - 20 \lg 4\pi),$$
(11)

where G_H and R_k are denoted as the transmit antenna gain and the UEs receive antenna gain, respectively. ϖ denotes the path loss coefficient, λ_F denotes the FSO channel wavelength, and $D_{H,k}$ represents the signal transmission distance from the HAP to the *k*th UEs.

In addition, $\mathbf{A}(\varphi_h, \theta_h)$ denotes the HAP-to-UEs array steering matrix, and $\varphi_h \in [0, 2/\pi)$ and $\theta_h \in [0, 2\pi)$ are the elevation and azimuth angles, respectively. $\mathbf{A}(\varphi_h, \theta_h)$ can be expressed as

$$\mathbf{A}(\varphi_h, \theta_h) = \mathbf{a}_x(\varphi_h, \theta_h) \otimes \mathbf{a}_y(\varphi_h, \theta_h), \tag{12}$$

where $\mathbf{a}_x(\varphi_h, \theta_h)$ and $\mathbf{a}_y(\varphi_h, \theta_h)$ denote the horizontal and vertical components of the antenna steering vector, respectively, which can be expressed as

$$\mathbf{a}_{x}(\varphi_{h},\theta_{h}) = \begin{bmatrix} e^{-j(2\pi d_{v}/\lambda)(1-(N_{1}+1)/2)\cos\varphi_{h}\cos\theta_{h}}, \\ \dots, e^{-j(2\pi d_{v}/\lambda)(N_{1}-(N_{1}+1)/2)\cos\varphi_{h}\cos\theta_{h}} \end{bmatrix}^{T}.$$
(13)

$$\mathbf{a}_{y}(\varphi_{h},\theta_{h}) = \begin{bmatrix} e^{-j(2\pi d_{h}/\lambda)(1-(N_{2}+1)/2)\cos\varphi_{h}\cos\theta_{h}}, \\ \dots, e^{-j(2\pi d_{h}/\lambda)(N_{2}-(N_{2}+1)/2)\cos\varphi_{h}\cos\theta_{h}} \end{bmatrix}^{T},$$
(14)

where d_v and d_h represent the physical spacings between adjacent elements of the antenna array in the *x*- and *y*-axis directions, respectively, and N_1 and N_2 denote the numbers of transmit antennas in the horizontal and vertical directions, respectively, which satisfy $N_1 \times N_2 = N_H$. By substituting Eqs. (13) and (14) into Eq. (12), we can derive the *i*th component as follows:

$$[\mathbf{A}(\varphi_h, \theta_h)]_i = [\mathbf{a}_x(\varphi_h, \theta_h)]_{n_1} [\mathbf{a}_y(\varphi_h, \theta_h)]_{n_2}$$

$$\stackrel{\Delta}{=} \exp\left(-j(i_1\cos\varphi_h\sin\theta_h + i_2\cos\varphi_h\cos\theta_h)\right),$$
(15)

where
$$i_1 = (2\pi d_v/\lambda)(n_1 - (N_1 + 1)/2)$$
, $n_1 = i/N_1$, $i_2 = (2\pi d_h/\lambda)(n_2 - (N_2 + 1)/2)$, and $n_2 = i - (n_1 - 1)N_1$.

Similarly, the HAP-to-RIS channel gain vector can be computed as

$$\mathbf{h}_{\mathrm{HR}} = \sqrt{\frac{\chi_{\mathrm{RD}}}{d_{\mathrm{HR}}^2}} \left(\sqrt{\frac{K_1}{1+K_1}} \bar{\mathbf{h}}_{\mathrm{HR}} + \sqrt{\frac{1}{1+K_1}} \tilde{\mathbf{h}}_{\mathrm{HR}} \right), \tag{16}$$

where χ_{RD} denotes the path loss value at a reference distance and the distance between the HAP and the RIS can be denoted as $d_{\text{HR}} = \sqrt{\|\mathbf{L}_H - \mathbf{q}[n]\|^2}$. $\mathbf{\tilde{h}}_{\text{HR}}$ denotes the deterministic LoS link component; $\mathbf{\tilde{h}}_{\text{HR}}$ denotes the random fast-fading non-line-of-sight (NLoS) component that are independently and identically distributed. K_1 is the Rician factor denoting the power ratio between $\mathbf{\bar{h}}_{\text{HR}}$ and $\mathbf{\tilde{h}}_{\text{HR}}$. Based on the corresponding antenna array response, $\mathbf{\bar{h}}_{\text{HR}}$ can be written as

$$\bar{\mathbf{h}}_{\mathrm{HR}} = \left[1, \dots, e^{-j2\pi \frac{d_r}{\lambda} \cos\phi}, \dots, e^{-j2\pi \frac{d_r}{\lambda} (M-1) \cos\phi}\right]^{\mathrm{T}},\tag{17}$$

where d_r denotes the spacing of each reflective element and $\cos \phi = \frac{x[n]-x_H}{\|\mathbf{q}[n]-\mathbf{L}_H\|}$ denotes the cosine of the angle of arrival (AoA) from the HAP to the RIS [27]. And, the channel vector gain from the RIS to the *k*th UEs can be written as

$$\mathbf{g}_{\mathrm{RG},k} = \sqrt{\frac{\chi_{\mathrm{RG}}}{d_{\mathrm{RG},k}}} \left(\sqrt{\frac{K_2}{K_2 + 1}} \bar{\mathbf{g}}_{\mathrm{RG},k} + \sqrt{\frac{1}{K_2 + 1}} \tilde{\mathbf{g}}_{\mathrm{RG},k} \right), \tag{18}$$

where β_{RG} denotes the associated path loss index and $d_{\text{RG},k} = \sqrt{\|\mathbf{q}[n] - \mathbf{b}_k\|^2}$ represents the distance from the RIS to the *k*th UEs [28]. Then, $\mathbf{\tilde{g}}_{\text{RG},k}$ can be computed as

$$\bar{\mathbf{g}}_{\mathrm{RG},k} = \left[1, \dots, e^{-j2\pi \frac{d_e}{\lambda} \cos \phi_{\mathrm{RG},k}}, \dots, e^{-j2\pi \frac{d_e}{\lambda} (M-1) \cos \phi_{\mathrm{RG},k}}\right]^T,$$
(19)

where $\cos \phi_{\text{RG},k} = \frac{x[n]-x_k}{\|\mathbf{q}[n]-\mathbf{b}_k\|}$ denotes the cosine of the angle of departure (AoD) from the RIS to the *k*th IoT device.

2.3 Energy consumption model

In the designed system model, the energy consumption model is mainly considered for UAVs because the communication-related energy consumption is significantly lower than the propulsion energy of UAVs. From [29], the flight energy of UAV can be written as

$$P_{\rm UAV} = P_B + P_C + P_D \tag{20}$$

According to [30, 31], the blade profile power consumption P_B and induced power consumption P_C can be expressed respectively as

$$P_B = P_h \left(1 + \frac{3V^2}{\Omega_B^2 r_R^2} \right),\tag{21}$$

Table 2 Abbreviations

Parameter	Specific meanings
Ω_B	UAV rotor angular velocity
r _R	Radius of the UAV rotor
$ ho_a$	Density of air
Sr	Solidity of UAV rotor
Ar	Disk area of UAV rotor
Vi	Propulsion speed in moving forward
df	Fuselage drag ratio
V	Current speed

$$P_C = P_S \left(\sqrt{1 + \frac{V^4}{4_{\nu_i}^4}} - \frac{V^2}{2_{\nu_i}^2} \right)^{1/2},\tag{22}$$

$$P_D = \frac{1}{2} d_f \rho_a s_r A_r V^3, \tag{23}$$

where P_h indicates the consumption of the leaf type in the hovering state, P_S is the consumption at the induction state, P_D denotes the parasitic power consumption, while Table 2 provides a summary of the physical interpretations of the parameters referred to in Eqs. (21)-(23). The overall energy consumption model can be computed as

$$E_{\rm UAV}(t) = P_{\rm UAV} t_{\rm fly},\tag{24}$$

where $t_{\rm fly}$ denotes the entire operating flight time of the UAV.

2.4 Problem formulation

_

The aim of this paper is to jointly optimize the active beamforming matrix w of the HAP, the passive beamforming reflection matrix Φ of the RIS, and the trajectory points \mathbf{Q} of the UAV to maximize the SER. Considering the signal transmission stage and the RIS phase modulation constraints, the problem of maximizing the long-term system efficiency can be expressed as follows:

$$\max_{\mathbf{Q},\mathbf{w},\mathbf{\Phi}} C(\mathbf{q},\mathbf{w},\mathbf{\Phi},\mathbf{h}_k,h_{\mathrm{SH}})$$
(25a)

$$s.tC1: E\left\{tr\left\{\mathbf{wx}(\mathbf{wx})^{H}\right\}\right\} \le P_t$$
(25b)

$$C2: x_{\min} \le x[n] \le x_{\max}, \tag{25c}$$

$$C3: y_{\min} \le y[n] \le y_{\max}, \tag{25d}$$

$$C4: \sum_{i \neq j}^{i=0} {}^{K}L_{\mathbf{q}_{i},\mathbf{q}_{j}} = 1, \forall \mathbf{q}_{i}, \mathbf{q}_{j} \in U,$$
(25e)

$$C5: |\exp(j\theta_m)| = 1, \forall m = 1, 2, ..., M,$$
(25f)

where *C*1 represents the transmit power constraint at the HAP; *C*2 and *C*3 correspond to the requirement for the UAV to complete its mission within a fixed area; *C*4 guarantees that the UAV will have only one path into and out of each hovering position, and *C*5 represents the phase-shift modulation constraints of the RIS elements [32].

The previous discussion highlights the problem as a constrained combinatorial optimization problem. As the number of UEs increases, the complexity of the optimization function also escalates, posing challenges for finding feasible solutions using traditional alternating optimization algorithms. To tackle this issue, a new approach is introduced in the paper. It utilizes the long short-term memory (LSTM) architecture within the framework of the double deep Q-network (DDQN) algorithm to obtain tractable solutions for these complex optimization problems. By incorporating LSTM with DDQN, the proposed method aims to address the growing complexity and provide efficient solutions to the optimization problem.

3 Joint optimization algorithm approach

In this section, the methodologies employed to address the joint optimization problems is represent. Firstly, we provide an introduction to the fundamental principles of LSTM neural network computation. This serves as a foundation for guiding the solution steps in our study. Next, we propose a novel approach based on the LSTM-DDQN framework. This approach leverages the information retention capability of LSTM networks to enhance the decision-making process of the DDQN algorithm. By integrating LSTM with DDQN, we aim to improve the effectiveness of the optimization process, enabling more accurate decision-making for joint optimization problems.

3.1 Preliminaries of LSTM

The LSTM network is a type of recurrent neural network known for its ability to retain and retrieve information from both short-term memory and long-term memory. This characteristic makes it highly suitable for tasks involving sequential data processing. Central to the LSTM architecture is the LSTM cell, which comprises four gates. These gates are governed by sigmoid activation functions and regulate the flow of information into the cell. The operations performed by these gates can be mathematically expressed as

$$f^{(t)} = \beta \left(W_f \cdot \left[h^{(t-1)}, x^{(t)} \right] + b_f \right), \tag{26a}$$

$$i^{(t)} = \beta \left(W_i \cdot \left[h^{(t-1)}, x^{(t)} \right] + b_i \right), \tag{26b}$$

$$o^{(t)} = \beta \left(W_o \cdot \left[h^{(t-1)}, x^{(t)} \right] + b_o \right), \tag{26c}$$

$$C^{(t)} = f^{(t)} \times C^{(t-1)}$$
(26d)

$$+i^{(t)} \times \tanh\left(W_C^{(t)} \cdot \left[h^{(t-1)}, x^{(t)}\right] + b_C\right),\tag{26e}$$

$$h^{(t)} = o^{(t)} \times \tanh C^{(t)},\tag{26f}$$

where $f^{(t)}$ represents the forget gate responsible for determining which information in the cell state to retain or discard. The input gate, denoted by $i^{(t)}$, decides which new information should be incorporated into the cell state based on the current input and previous hidden state. The output gate, $o^{(t)}$, controls the extent to which the information in the cell state is utilized to generate the current output. The weights associated with each gate are represented by W_f , W_i , and W_o while b_f , b_i , and b_o are the corresponding bias terms. The cell state and hidden state at the previous time step are denoted as $C^{(t-1)}$ and $h^{(t-1)}$. The activation functions $\beta(\cdot)$ and $\tanh(\cdot)$ are applied to certain intermediate results.

While the LSTM gates play a crucial role in information flow regulation, the LSTM cell also includes control units that enhance its adaptability and flexibility for various tasks. These control units enable the LSTM network to adjust the flow of information based on the specific requirements of the task at hand. This capability makes LSTM networks particularly useful for processing and forecasting time series data, including significant events. After passing through the gates and undergoing correlation operations, the input information is modified, resulting in the updated unit state $C^{(t)}$ and hidden state $h^{(t)}$. These states are continuously updated by the gating units, which selectively control the flow of information into and out of the LSTM cell. By incorporating an LSTM network, a UAV (unmanned aerial vehicle) can learn from its previous experiences and make informed decisions based on the accumulated knowledge. In our approach, we aim to incorporate the LSTM architecture into the construction of deep convolutional neural network models to enable long-term memory of the explored environmental state information. Prior to integrating the LSTM architecture, the network model can be represented as follows:

$$\begin{cases} X^{(t)} = \text{EncoderProcess}(x^{(t)}), \\ X^{(t)}, h^{(t+1)}, C^{(t+1)} = \text{LSTM-DDQN}(X^{(t)}, h^{(t)}, C^{(t)}), \\ a^{(t)} = \text{DecoderProcess}(X^{(t)}), \end{cases}$$
(27)

where EncoderProcess(·) denotes the encoding function. On the other hand, the decoding process is performed by multilayer fully connected network denoted as DecoderProcess(·). Besides, we use $(x^{(t)})$ to represent the input observation data, and $X^{(t)}$ is an implicit representation of the observed data encoded using a neural network. This formulation allows us to capture and utilize the underlying patterns and features of the observed data in an efficient manner. By incorporating both the encoding and decoding processes, our model can effectively learn and generate meaningful outputs based on the input observations.

3.2 Algorithm process

This article proposes an improved DDQN algorithm that integrates the LSTM and DDQN architectures. The network structure is shown in Fig. 2. The state historical information and channel information obtained by the UAV during exploration are used



Fig. 2 The improved network structure of LSTM-DDQN algorithm

as inputs. The state features are extracted through three convolutional layers and then passed to an LSTM-DNN layer, which is used for long-term memory and storage of the explored environmental states. The LSTM-DNN layer is followed by a second fully connected layer, which outputs Q-values as a basis for choosing among the possible actions.

With the introduction of an LSTM layer into the neural network structure, useful historical state information can be stored in long-term memory, allowing the agent to make informed decisions based on past experiences when exploring unknown environments. With this neural network, the newly transformed state space $s_n^{(t)}$ can be written as a function of the previous state and the current observation. The agent gradually improves its policy by adapting its behavior based on the feedback received from the environment. By leveraging the MDP framework, the DRL approach allows the agent to learn from its interactions with the environment over time and maximize its cumulative reward [33].

1) State space: In deep reinforcement learning (DRL), the state space refers to the collection of all possible states that describe the environment. A state represents the environment at a specific moment and encompasses all pertinent information required for an intelligent agent to make decisions based on the current circumstances. In this paper, we focus on a discrete state space, where the set of possible states is finite, allowing the intelligent agent to choose from a limited number of discrete states. The design of the state space plays a crucial role in the success of reinforcement learning. It should effectively capture the essential characteristics of the environment, encompassing relevant input and target variables. This enables the intelligent agent to accurately perceive and adapt to the dynamic changes within the environment, optimizing the effectiveness of its strategies. In this paper, the state space consists of the current location of the UAV, the channel characteristics, the energy consumed by the UAV, and the action taken in the (t - 1)th time step.

2) Action space: In DRL-based optimization framework, the action space represents the set of all possible actions that an intelligent agent can take. These actions allow the agent to interact with the environment and influence its state by selecting different actions. In each time step, the agent is required to select an action from the available action space to execute and advance toward its optimization goal. This action selection process enables the agent to gradually move closer to achieving its desired outcome. Through the utilization of techniques like deep neural networks, the agent can learn the mapping between states and actions, enabling it to make optimal decisions. The design of the action space is intricately connected to the agent's strategy and decision-making process, significantly impacting the achievement of reinforcement learning objectives. The action space consists of three parts, the change in the phase shift of each reflective element, the change in the values of the transmit beamforming matrix, and the change in the UAV's position. By considering the existing conditions, the agent makes informed decisions to select the most appropriate action that aligns with its goals and objectives.

3) Reward: The reward function refers to the spectrum of feedback signals received by an intelligent agent based on its actions. Rewards serve as a measure to evaluate the quality of the agent's behavior and act as a feedback mechanism to indicate whether it is progressing toward its optimization goal. They can be represented as real numbers, reflecting the desirability of the agent's actions. Positive rewards typically signify favorable feedback, encouraging the agent to reinforce and increase such behavior. Conversely, negative rewards indicate unfavorable feedback, prompting the agent to avoid or diminish those actions. Zero rewards may represent neutral feedback without significant influence on the agent decision-making process. In this paper, we consider reward functions with respect to the SER [34], R(t), and includes the transmit beamforming matrix, the RIS phase shifts, the UAV position, and the direction of motion selection for UAV, denoted as

$$r^{(t)} = \frac{R(t)}{E_{\rm UAV}(t)},$$
 (28)

where the environment provides feedback to the agent in the form of a reward/penalty function. In this paper, we propose a reinforcement learning framework to enhance the overall energy efficiency (EE) of the system by incentivizing the agent to select actions that improve performance. When the agent chooses actions that lead to improved EE, it receives positive rewards from the environment. This reinforcement mechanism reinforces the agent's behavior, encouraging it to repeat similar actions in the future. To address the optimization problem using the LSTM-DDQN algorithm, we follow several steps. Firstly, we initialize the hyperparameters of the LSTM network and the environmental parameters in the communication scenario. The initial coordinates of the UAV and RIS phase shifts are used to calculate the channel coefficients and determine the user rate of the current state. Next, we feed the state information into the LSTM network and employ selection operations to obtain the next state. To further improve the algorithm's performance and efficiency in solving optimization problems that involve transmit beamforming, RIS phase shift, and UAV trajectory design, we integrate the prioritized experience replay strategy with the LSTM architecture and DDQN algorithm.

4 Simulation experiment and result analysis

In this section, we present the performance evaluation of our proposed system with an advanced DRL-based LSTM-DDQN framework and analyze the transmission model under various parameter settings to assess its effectiveness. The main system model other parameters used for simulation are provided in Table 3. To compare the performance of our proposed algorithm with other simulation methods, we have designed four benchmark scenarios that demonstrate its superiority in the considered system model. Specifically, these benchmark scenarios are carefully selected to reflect different system

Table 3 Simulation parameter settings

	Value
Satellite type	GEO
FSO channel bandwidth	$B_F = 1280 \text{ MHz}$
FSO channel wavelength	$\lambda_F = 1550 \mathrm{nm}$
Carrier frequency band of RF links	2 GHz
RF channel bandwidth	$B_R = 1 \text{MHz}$
Maximum height of UAV	$z[n]_{max} = 5000 \mathrm{m}$
IoT devices area	1000 m × 1000 m
DNN parameter	
Number of time slots	100
Reward discount rate	0.99
Number of episodes	10,000
Number of training rounds per step	2000
Number of steps per episode	40,000
Number of experiences in a minibatch	64

setups, including varying numbers of users and RIS elements, different channel conditions, and diverse transmission distances, which can be expressed as

Benchmark 1- RIS-NOMA LSTM-DDQN scheme: We design the proposed LSTM-DDQN algorithm with prioritized experience replay. This modification helps us tackle the problem more effectively by optimizing the values of \mathbf{Q} , Φ , and \mathbf{w} while ensuring that the capacity constraint is satisfied.

Benchmark 2- RIS-NOMA DDQN scheme: Unlike Benchmark 1, our proposed approach deviates from relying on historical information and instead promotes the use of a traditional DDQN scheme. In this scheme, the agent decision-making process is independent of any influence from past experiences or historical data. By adopting a traditional DDQN scheme, our proposal aims to simplify the decision-making process and reduce computational complexity. This approach allows the agent to focus solely on the current state of the system and make decisions based on immediate information.

Benchmark 3- RIS-OMA LSTM-DDQN scheme: In contrast to other benchmarks, our proposal recommends the utilization of the proposed LSTM-DDQN scheme, which incorporates the OMA-enabled transmission strategy.

Benchmark 4- RIS-OMA DDQN scheme: In contrast to Benchmark 3, our proposal suggests the utilization of both the traditional DDQN scheme and the OMA-enabled transmission scheme. This approach combines the strengths of the two methodologies to show the enhanced system performance.

Figure 3 shows the relationship between the reward function and the number of reflection elements during the traversal process of the optimization schemes using different algorithms, where the meaning of the reward function can be interpreted as the energy efficiency, which is closely related to the energy consumption during the whole communication process. In this scenario, a UAV is launched from a fixed position, and an algorithm is employed to determine the optimal position at each operational step. Throughout this process, the UAV dynamically adjusts its active transmit beamforming



Fig. 3 Comparison of rewards under different optimization algorithms

matrix and the phase shift of the RIS. As depicted in Fig. 3, the number of convergence steps for the LSTM-DDQN-based algorithm closely matches that of the conventional DDQN algorithm. It only takes approximately 320 steps, validating the effectiveness of the state space and rewards designed in this study, as well as the proposed algorithm's performance. Moreover, it highlights the potential of integrating LSTM into the DDQN framework, demonstrating promising outcomes.

From Fig. 4, we can conclude that with the increase of RIS reflection elements, the energy efficiency after optimization using the LSTM-DDQN algorithm gets significantly increased. At M=16, the energy efficiency after optimization using LSTM-DDQN is increased by 7% compared to the traditional DDQN algorithm and at M = 128, the energy efficiency after optimization using LSTM-DDQN is increased by 23% compared to the traditional DDQN algorithm. By incorporating LSTM networks into the DDQN framework, temporal correlations can be effectively captured [35]. This is achieved by utilizing LSTM's ability to learn dynamic patterns of sequence data through the inclusion of outdated but useful information as input sequences. Moreover, the LSTM-based DDQN framework benefits from memory units and gating mechanisms, which enable it to handle long-term dependencies present in the sequence data. However, when transitioning from M = 16 to M = 32 RIS elements, the performance gain resulting from the increased number of components is not significantly apparent. The substantial performance enhancement observed with a higher number of RIS elements can be attributed to several factors. Firstly, RIS can effectively enhance the signal strength and quality by reflecting and manipulating the incident signals. By adjusting the phase and amplitude of the reflected signals, RIS can optimize the channel conditions and mitigate the effects of fading and interference. Secondly, RIS enables precise beamforming and waveform shaping. This increases the signal power in specific areas or toward specific users, resulting in improved coverage, reduced interference, and enhanced link quality. Furthermore,



Fig. 4 Comparison of system energy efficiency with different RIS reflective elements

RIS can enable enhanced spatial multiplexing and diversity gain. By exploiting the reflective properties of RIS, it becomes possible to create multiple signal paths between the transmitter and the receiver. Additionally, RIS can enhance diversity gain by introducing additional signal paths that mitigate the effects of fading and improve signal reliability.

5 Conclusions

In this paper, we introduced a novel architecture for downlink massive access RIS-UAV relay-assisted in hybrid FSO/RF-based ISATRNs. A secure signal transmission model was first established by defining a target optimization problem based on different transmission modes at various stages. Secondly, we leveraged the performance of DRL technology for its model-free nature. Then, we further decomposed the optimization problem into subproblems, including trajectory optimization for the UAV, active beamforming matrix, and RIS phase shift. To address these subproblems under constraints, we proposed a novel DRL-based LSTM-DDQN algorithm framework to supplement the current state with historical information. The proposed LSTM-DDQN algorithm had strong scalability to a certain extent with prioritized experience replay due to high dimensional state space and partial observability. Finally, numerical simulation results demonstrated the superiority of the LSTM-DDQN algorithm and verified the impact of the number of RIS reflection elements, and the transmit power level on SER.

Acknowledgements

The authors would like to extend their gratitude to the anonymous reviewers for their valuable and constructive comments, which have largely improved and clarified this paper.

Author contributions

JL, HX, MW, FW, TG, and FZ conceived of and designed the experiments. JL and HX performed the experiments. JL, MW, and FW analyzed the data. TG contributed analysis tools; JL, HX, MW, and FZ wrote the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the Soft Science Research Project of Henan Province under Grant 222400410137, and in part by the Key R &D projects in the autonomous region under Grant 2020B02018-2, and the Natural Science Foundation of Universities of Anhui Province under Grant KJ2020A0694.

Availability of data and materials

The raw/processed data required to reproduce the above findings cannot be shared at this time as the data also forms part of an ongoing study.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 11 August 2023 Accepted: 11 October 2023 Published online: 27 November 2023

References

- N. Saeed, H. Almorad, H. Dahrouj, T.Y. Al-Naffouri, J.S. Shamma, M.-S. Alouini, Point-to-point communication in integrated satellite-aerial 6G networks: state-of-the-art and future challenges. IEEE Open J. Commun. Soc. 2(2), 1505–1525 (2021)
- X. Zhu, C. Jiang, Integrated satellite-terrestrial networks toward 6G: architectures, applications, and challenges. IEEE Internet Things J. 9(1), 437–461 (2022)
- K. An, M. Lin, J. Ouyang, W.P. Zhu, Secure transmission in cognitive satellite terrestrial networks. IEEE J. Sel. Areas Commun. 34(11), 3025–3037 (2016)
- 4. Z. Lin, M. Lin, T. de Cola, J.-B. Wang, W.-P. Zhu, J. Cheng, Supporting IoT with rate-splitting multiple access in satellite and aerial-integrated networks. IEEE Internet Things J. 8(14), 11123–11134 (2021)
- F. Zhou, X. Li, M. Alazab, R.H. Jhaveri, K. Guo, Secrecy performance for RIS-based integrated satellite vehicle networks with a UAV relay and MRC eavesdropping. IEEE Trans. Intell. Veh. 8(2), 1676–1685 (2023)
- X. Zhang, D. Guo, K. An, G. Zheng, S. Chatzinotas, B. Zhang, Auction-based multichannel cooperative spectrum sharing in hybrid satellite–terrestrial IoT networks. IEEE Internet Things J. 8(8), 7009–7023 (2021)
- M.A. Khalighi, M. Uysal, Survey on free space optical communication: a communication theory perspective. IEEE Commun. Surv. Tutor. 16(4), 2231–2258 (2014)
- G. Xu, N. Zhang, M. Xu, Z. Xu, Q. Zhang, Z. Song, Outage probability and average BER of UAV-assisted dual-hop FSO communication with amplify-and-forward relaying. IEEE Trans. Veh. Technol. 72(7), 8287–8302 (2023)
- G. Pan, J. Ye, J. An, M.-S. Alouini, Full-duplex enabled intelligent reflecting surface systems: opportunities and challenges. IEEE Wirel. Commun. 28(3), 122–129 (2021)
- L. Lv, Q. Wu, Z. Li, Z. Ding, N. Al-Dhahir, J. Chen, Covert communication in intelligent reflecting surface-assisted NOMA systems: design, analysis, and optimization. IEEE Trans. Wirel. Commun. 21(3), 1735–1750 (2022)
- 11. T. Wang, F. Fang, Z. Ding, An SCA and relaxation based energy efficiency optimization for multi-user RIS-assisted NOMA networks. IEEE Trans. Veh. Technol. **71**(6), 6843–6847 (2022)
- 12. Z. Lin, M. Lin, J.-B. Wang, T. de Cola, J. Wang, Joint beamforming and power allocation for satellite–terrestrial integrated networks with non-orthogonal multiple access. IEEE J. Sel. Top. Signal Process. **13**(3), 657–670 (2019)
- Z. Na, Y. Liu, J. Shi, C. Liu, Z. Gao, UAV-supported clustered NOMA for 6G-enabled Internet of Things: trajectory planning and resource allocation. IEEE Internet Things J. 8(20), 15041–15048 (2021)
- S. Yang, J. Zhang, W. Xia, Y. Ren, H. Yin, H. Zhu, A unified framework for distributed RIS-aided downlink systems between MIMO-NOMA and MIMO-SDMA. IEEE Trans. Commun. 70(9), 6310–6324 (2022)
- X. Pang et al., When UAV meets IRS: expanding air-ground networks via passive reflection. IEEE Wirel. Commun. 28(5), 164–170 (2021)
- L. Lv, Z. Ding, J. Chen, N. Al-Dhahir, Design of secure NOMA against full-duplex proactive eavesdropping. IEEE Wirel. Commun. Lett. 8(4), 1090–1094 (2019)
- K. Guo, M. Wu, X. Li, H. Song, N. Kumar, Deep reinforcement learning and NOMA-based multi-objective RIS-assisted IS-UAV-TNs: trajectory optimization and beamforming design. IEEE Trans. Intell. Transp. Syst. 24(9), 10197–10210 (2023)
- K. Li, C. Huang, Y. Gong, G. Chen, Double deep learning for joint phase-shift and beamforming based on cascaded channels in RIS-assisted MIMO networks. IEEE Wirel. Commun. Lett. 12(4), 659–663 (2023)
- P. Chen, W. Huang, X. Li, S. Jin, Deep reinforcement learning based power minimization for RIS-assisted MISO-OFDM systems. China Commun. 20(4), 259–269 (2023)
- 20. Z. Peng, Z. Zhang, L. Kong, C. Pan, L. Li, J. Wang, Deep reinforcement learning for RIS-aided multiuser full-duplex secure communications with hardware impairments. IEEE Internet Things J. **9**(21), 21121–21135 (2023)
- X. Li et al., Physical-layer authentication for ambient backscatter-aided NOMA symbiotic systems. IEEE Trans. Commun. 71(4), 2288–2303 (2023)
- H. Niu, Z. Chu, F. Zhou, P. Xiao, N. Al-Dhahir, Weighted sum rate optimization for STAR-RIS-assisted MIMO system. IEEE Trans. Veh. Technol. 71(2), 2122–2127 (2022)
- 23. K. Guo et al., Physical layer security for multiuser satellite communication systems with threshold-based scheduling scheme. IEEE Trans. Veh. Technol. **69**(5), 5129–5141 (2020)
- C. Huang, A. Zappone, G.C. Alexandropoulos, M. Debbah, C. Yuen, Reconfigurable intelligent surfaces for energy efficiency in wireless communication. IEEE Trans. Wirel. Commun. 18(8), 4157–4170 (2019)

- C. Gong, X. Yue, X. Wang, X. Dai, R. Zou, M. Essaaidi, Intelligent reflecting surface aided secure communications for NOMA networks. IEEE Trans. Veh. Technol. 71(3), 2761–2773 (2022)
- M.A. Al-Habash, Mathematical model for the irradiance probability density function of a laser beam propagating through turbulent media. Opt. Eng. 40(8), 1554–1562 (2001)
- J. Gao, Y. Wu, S. Shao, W. Yang, H.V. Poor, Energy efficiency of massive random access in MIMO quasi-static Rayleigh fading channels with finite blocklength. IEEE Trans. Inf. Theory 69(3), 1618–1657 (2023)
- Z. Jia, M. Sheng, J. Li, D. Niyato, Z. Han, LEO-satellite-assisted UAV: joint trajectory and data collection for internet of remote things in 6G aerial access networks. IEEE Internet Things J. 8(12), 9814–9826 (2021)
- H.-T. Ye, X. Kang, J. Joung, Y.-C. Liang, Optimization for full-duplex rotary-wing UAV enabled wireless-powered IoT networks. IEEE Trans. Wirel. Commun. 19(7), 5057–5072 (2020)
- Y. Zeng, J. Xu, R. Zhang, Energy minimization for wireless communication with rotary-wing UAV. IEEE Trans. Wirel. Commun. 18(4), 2329–2345 (2013)
- H. Zhang, M. Huang, H. Zhou, X. Wang, N. Wang, K. Long, Capacity maximization in RIS-UAV networks: a DDQNbased trajectory and phase shift optimization approach. IEEE Trans. Wirel. Commun. 22(4), 2583–2591 (2023)
- X. Pang, N. Zhao, J. Tang, C. Wu, D. Niyato, K.-K. Wong, IRS-assisted secure UAV transmission via joint trajectory and beamforming design. IEEE Trans. Commun. 70(2), 1140–1152 (2022)
- N. Zhao, Z. Ye, Y. Pei, Y.-C. Liang, D. Niyato, Multi-agent deep reinforcement learning for task offloading in UAVassisted mobile edge computing. IEEE Trans. Wirel. Commun. 21(9), 6949–6960 (2022)
- 34. X. Liu, Y. Liu, Y. Chen, H.V. Poor, RIS enhanced massive non-orthogonal multiple access networks: deployment and passive beamforming design. IEEE J. Sel. Areas Commun. **39**(4), 1057–1071 (2021)
- J. Xu, B. Ai, Deep reinforcement learning for handover-aware MPTCP congestion control in space-ground integrated network of railways. IEEE Wirel. Commun. 28(6), 200–207 (2021)

Publisher's Note

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

Submit your manuscript to a SpringerOpen[™] journal and benefit from:

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

Submit your next manuscript at > springeropen.com