Joint Trajectory and Passive Beamforming Design for Intelligent Reflecting Surface-Aided UAV Communications: A Deep Reinforcement Learning Approach

Liang Wang, Kezhi Wang, Cunhua Pan, and Nauman Aslam

Abstract—In this paper, the intelligent reflecting surface (IRS)-aided unmanned aerial vehicle (UAV) communication system is studied, where the UAV is deployed to serve the user equipment (UE) with the assistance of multiple IRSs mounted on several buildings to enhance the communication quality between UAV and UE. We aim to maximize the energy efficiency of the system, including the data rate of UE and the energy consumption of UAV via jointly optimizing the UAV’s trajectory and the phase shifts of reflecting elements of IRS, when the UE moves and the selection of IRSs is considered for the energy saving purpose. Since the system is complex and the environment is dynamic, it is challenging to derive low-complexity algorithms by using conventional optimization methods. To address this issue, we first propose a deep Q-network (DQN)-based algorithm by discretizing the trajectory, which has the advantage of training time. Furthermore, we propose a deep deterministic policy gradient (DDPG)-based algorithm to tackle the case with continuous trajectory for achieving better performance. The experimental results show that the proposed algorithms achieve considerable performance compared to other traditional solutions.

Index Terms—Deep reinforcement learning, UAV communications, intelligent reflecting surface

1 INTRODUCTION

It is widely envisioned that the fifth-generation (5G) wireless networks and beyond will achieve 1000-fold increase in network capacity, accommodate about 10 billion devices and support a number of emerging applications such as virtual reality (VR) services. To satisfy this ever-increasing demand, unmanned aerial vehicle (UAV) has been applied and regarded as one of the most promising technologies to achieve these ambitious goals. Compared to the traditional communication systems that utilize the terrestrial fixed base stations, UAV-aided communication systems are more cost-effective and likely to achieve better quality of service (QoS) due to its appealing properties of flexible deployment, fully controllable mobility and low cost. In fact, with the assistance of UAVs, the system performance (e.g., data rate and latency) can be significantly enhanced by establishing the line-of-sight (LoS) communication links between UAVs and user equipments (UEs).

In addition, to further improve the channel quality, adaptive communications can be designed through the mobility/deployment control of the UAV systems. For example, in [1], Jiang et al. proposed a heterogeneous mobile edge computing (MEC) framework, where ground stations (GSs), ground vehicles (GVs) and UAVs are deployed for providing computing, communication and caching (3C) resources at the network edge. In [2], Yang et al. investigated the weighted-sum cost minimization problem in a hierarchical machine learning (ML) tasks distribution (HMTD) framework and they optimized the offloading strategy, including the binary offloading and partial offloading between the UAV and target. In [3], Hourani et al. proposed an analytical approach for optimizing the altitude of UAV for the purpose of maximizing the radio coverage on the ground. In [4], the authors considered the scenario of UAVs in an orthogonal frequency division multiple access (OFDMA) system and they proposed an iterative block coordinate descent approach for optimizing the UAV’s trajectory and resource allocation, aiming to maximize the minimum average throughput of UEs. The optimization problem of UAV placement and transmit power in UAV-aided relay systems was studied in [5], where Ren et al. proposed a low-complexity iterative algorithm to solve the problem both in the free-space channel and three-dimensional channel scenarios. In [6], to minimize the energy consumption of UAV, Zeng et al. formulated a travelling sale problem and proposed an efficient algorithm to optimize the UAV’s trajectory, including the hovering locations and duration. In [7], a multi-UAV-assisted communication system was studied. The authors proposed a energy-efficient distributed MCS...
In [22], the authors investigated the Holographic Reflecting Surface (RIS), or reconfigurable intelligent surfaces (RIS) [18], [19], has been proposed and received considerable attention in both academia and industry. In [18], [19], Yang et al. studied the power minimization problem in a multi-UAV-enabled MEC system, where the user association, power control, computation capacity allocation and location planning were optimized. In [19], Wang et al. investigated energy minimization problem in the multi-UAV assisted MEC system, where they proposed a multi-agent deep reinforcement learning approach for optimizing the trajectories of UAVs. In [20], the authors proposed a convex optimization based trajectory (CAT) and deep Reinforcement learning based trajectory (RAT) algorithms for optimizing the user association, resource allocation and the trajectory of UAVs, aiming at minimizing the energy consumption of UEs. In [21], Huang et al. investigated the device-to-device (D2D) rate maximization problem in UAV-aided wireless communication systems, where they proposed an iterative algorithm for optimizing the UAV flying altitude, location and the bandwidth allocation, which proved that the altitude of the UAV is vital for improving the system performance. In [22], Liu et al. introduced a distributed mobile crowed sensing platform, where multiple UAVs are deployed as mobile terminals for collecting data. They proposed a DRL-based approach for navigating a group of UAVs in order to maximize the collected data, the geographical fairness, and the energy efficiency of UAVs. In [23], Xu et al. studied the problem of maximizing the energy harvested at all energy receivers in a UAV-enabled wireless power transfer system, in which they first proposed an algorithm based on Lagrange dual method for optimizing UAVs’ trajectory in an ideal case. Then, they proposed a new successive hover-and-fly algorithm based on convex programming optimization for trajectory design for the general case.

However, in the crowded area, the communication signals between UAV and UE may be blocked by high buildings or other constructions. Thanks to the development of meta-materials or meta-surfaces [16], [17], intelligent reflecting surface (IRS), or reconfigurable intelligent surfaces (RIS) [18], [19] has been proposed and received considerable attention in both academia and industry. In general, the IRS consists of an array of low-cost and passive reflecting elements, each of which is able to reflect the incident signals by smartly adjusting the phase shift, which has the potential to improve the achievable data rate [20]. Furthermore, since the reflecting elements of the IRS can be passive, the IRS is more energy-efficient than traditional relay-aided communication techniques, such as [21].

Due to the above advantages, the IRS has been extensively investigated in various wireless communication systems. In [24], the authors investigated the Holographic Multiple Input Multiple Output Surface (HMIMOS) architecture and analyzed its opportunities and challenges in 6G wireless networks. In [25], an IRS-enhanced MISO wireless system was studied, and the authors proposed a semidefinite relaxation (SDR) based algorithm for optimizing the active and passive beamforming, aiming to maximize the overall received signal power at the user. In [26], Yang et al. studied a realistic IRS-enhanced OFDM system, where the frequency-selective channels were considered, and the passive array reflecting coefficients were optimized for maximizing the achievable data rate of the user. In order to enhance the physical layer security of IRS-aided communication systems, Yu et al. [27] jointly optimized the beamforming at the transmitter and the phase shifts of the IRS, maximizing the physical layer security data rate. For multi-cast scenarios, the authors in [28] investigated the downlink IRS-aided multigroup multicast communication system, where the IRS can be deployed to enhance the worst-case user channel condition. In [29], Pan et al. studied the weighted sum rate (WSR) maximization problem for an IRS-assisted multicell MIMO communication system, and the authors proposed a pair of algorithms named Majorization-Minimization (MM) and Complex Circle Manifold (CCM) for optimizing the phase shifts of the IRS. The simulation results in [20] showed that the IRS is very effective in mitigating the cell-edge interference. Additionally, the authors in [21] considered to deploy an IRS in a simultaneous wireless information and power transfer (SWIPT) system to enhance both the energy harvesting and data rate performance. In [30], the IRS was shown to be beneficial in reducing the latency of the mobile edge computing system. In [31], the authors studied the achievable rate problem in an IRS-aided wireless system, and they optimized the transmit beamforming and the IRS reflect beamforming through the alternating optimization (AO) based technique. In [32], the authors studied the resource allocation for a point-to-point IRS-aided MIMO communication system when taking into account the channel estimation and channel feedback overhead. In [33], Huang et al. proposed a DRL-based algorithm to optimize the design of beamforming matrix and phase shift matrix in RIS-based multi-user MISO system.

Against the above background, we study an IRS-aided UAV system where the UAV is deployed to provide communication services to the ground UE. To enhance the channel condition between UAV and UE, which may be blocked by some obstacles such as high buildings, the IRS may be mounted on the exterior wall of the buildings. We aim to maximize the energy efficiency of UAV, including the data rate of UE and the energy consumption of UAV via jointly optimizing the UAV’s trajectory, the phase shifts of the reflecting elements of IRS, while UE moves. To address this problem, first, we propose a deep Q-network (DQN)-based algorithm by discretizing the trajectory for the easy deployment. Then, we propose a deep deterministic policy gradient (DDPG)-based algorithm to tackle the continuous situation for better performance. The experiment verifies that the proposed algorithms achieve better performance compared to benchmark solutions.
TABLE 1
Main Notations

| Notation | Definition |
|----------|------------|
| $k, K, \mathcal{K}$ | the index, the number, and the set of IRSs |
| $Z_{\text{min}}, Z_{\text{max}}$ | the minimal, maximal of flying altitude of UAV |
| $x_{\text{max}}, y_{\text{max}}$ | side length of target area |
| $t, T, T'$ | the index, the number, and the set of TSs |
| $M_x, M_y$ | number of reflecting elements of IRSs |
| $x_{\text{max}}, y_{\text{max}}, z_{\text{max}}$ | maximal flying distances of UAV |
| $a_k^x, a_k^y, a_k^z$ | flying distances of UAV in $t$ |
| $\mathbf{x}_0, y_0, z_0$ | initial coordinate of UAV |
| $V_r$ | the mean rotor induced velocity when hovering |
| $d_0$ | the main body drag ratio |
| $\rho$ | air density |
| $z$ | the rotor solidity |
| $G$ | rotor disc area |
| $t_d$ | time duration of TS |
| $[x_k, y_k, z_k]$ | coordinate of IRS $k$ |
| $[x_t, y_t, z_t]$ | coordinate of UE in $TS_t$ |
| $[x_k^u, y_k^u, z_k^u]$ | coordinate of UAV in $TS_t$ |
| $d_{k,t}^\text{UL}$ | distance between UAV and IRS $k$ in $TS_t$ |
| $d_{k,t}^\text{UL}$ | distance between UE and IRS $k$ in $TS_t$ |
| $h_{k,t}^\text{UL}$ | channel gain of UAV-IRS $k$ link in $TS_t$ |
| $\mu$ | path loss at reference distance $1\text{m}$ |
| $\alpha_{\text{EE}}$ | path loss exponent |
| $f, c$ | carrier frequency, speed of light |
| $h_{k,t}^\text{IE}$ | channel gain of IRS $k$ - UE link in $TS_t$ |
| $P, \sigma^2, B$ | transmission power, noise power, bandwidth |
| $\Theta_{k,t}$ | phase shift matrix of IRS $k$ in $TS_t$ |
| $R_{k,t}$ | data rate of UAV-IRS $k$-UE link in $TS_t$ |

Other Notations: In this paper, $\mathbb{C}^{M \times 1}$ denotes the set of $M \times 1$ complex vectors. $\text{diag}(\cdot)$ denotes the diagonalization operation. $(\cdot)^T$ denotes the transpose operation. $E[\cdot]$ denotes the expectation operation.

2 RELATED WORK AND BACKGROUND

2.1 IRS-Aided UAV Communications

Most recently, the integration of IRS in UAV-aided communication systems has become a hot research topic. For example, in [33], the authors considered a downlink transmission system, consisting of a rotary-wing UAV, a ground user and an IRS. In this work, the authors proposed a successive convex approximation (SCA) based algorithm to optimize the UAV’s trajectory and passive beamforming of the IRS. In [34], the potential of IRS in UAV-assisted communication systems was investigated. The authors concluded that the deployment of IRS is capable of achieving significant performance gain in UAV-assisted cellular networks. Some other benefits of IRS-assisted system can also be found in the literature. Most of the existing algorithms are based on convex optimization theory, which may achieve suboptimal performance and is time-consuming due to the fact that a number of iterations are required for the convergence of the algorithm. Their complexity may increase with the number of reflecting elements.

2.2 DRL Background

Thanks to the advances in the field of machine learning, most of sophisticated optimization problems may be solved efficiently and in real time. As a branch of machine learning algorithms, reinforcement learning (RL) is viewed as a useful approach for tackling complicated control tasks, such as robotics and games. In [35], Sutton et al. proposed a widely used model-free RL algorithm named Q-learning, where some fundamental knowledge, such as agent, environment, state, action, reward and Q-value were introduced. In addition, another mechanism named Q-table was employed in Q-learning. However, as the size of Q-table is finite, Q-learning may only handle control problems in discrete state and action spaces. As an extension of Q-learning, Mnih et al. [36] proposed the deep Q-network (DQN) algorithm, which combines RL and the powerful deep neural network (DNN). Additionally, two techniques named experience replay and target network were integrated. The experimental results proved that DQN is capable of achieving enhanced performance in the challenging Atari 2600 games. In DQN, the Q-table is replaced by the DNN, as DQN can handle the control problem with infinite state spaces. However, the action space of DQN is still discrete. Inspired by DQN, Silver et al. proposed a deep deterministic policy gradient (DDPG) [37] algorithm based on the actor-critic [38] method, which is able to be applied to continuous action spaces. Although some researchers have started to apply the DRL in the IRS or IRS-assisted UAV communications, most of the work did not consider the selection of IRS and the movement of UE. In this paper, DDPG and DQN will be applied in IRS-aided UAV system, where the selection of IRS and the movement of UE will also be considered.

3 SYSTEM MODEL

Assume that there is one rotary UAV, $K$ IRSs mounted on $K$ buildings, respectively and one moving UE to be served, as shown in Fig. 1. Note that the UE can also be other moving object, like the autonomous vehicle. Also, assume that the UE is located in the crowded area where it suffers from severe path loss and high attenuation, caused by high buildings and trees. Thus, the direct link between UAV and UE is
not considered. IRSs are deployed for enhancing the communication quality of UE. The UAV flies within a particular altitude ranging from $[z_{\text{min}}, z_{\text{max}}]$ over a rectangular target area with side lengths $X_{\text{max}}$ and $Y_{\text{max}}$ for a certain number of time slots (TSs) $T$, each of which has $t_d$ time duration. For simplicity, we denote the set of IRSs as $\mathcal{K} \triangleq \{k = 1, 2, \ldots, K\}$ and the set of TSs as $\mathcal{T} \triangleq \{1, 2, \ldots, T\}$. Additionally, each of IRSs is equipped with an uniform rectangular array (URA) with $M_r \times M_t$ reflecting elements, which could boost the useful signal power by adjusting the phase shifts of the reflecting elements.

### 3.1 UAV Model

In this subsection, we describe the UAV model with Cartesian coordinate system. Specifically, in each of TS, the UAV moves with a flying action determined by two horizontal distances $a_i^x \in [-X_{\text{max}}, X_{\text{max}}]$, $a_i^y \in [-Y_{\text{max}}, Y_{\text{max}}]$ and a vertical distance $a_i^z \in [-Z_{\text{max}}, Z_{\text{max}}]$. Thus, given the initial coordinate of the UAV, which is $[x_0^u, y_0^u, z_0^u]$, the coordinate of the UAV in TS $t$ is expressed as $[x_t^u, y_t^u, z_t^u]$, where $x_t^u = x_0^u + \sum_{i=1}^{t} a_i^x$, $y_t^u = y_0^u + \sum_{i=1}^{t} a_i^y$, and $z_t^u = z_0^u + \sum_{i=1}^{t} a_i^z$. Note that as the UAV may not go beyond the border of the targeted area, we have the following constraints

$$0 \leq x_t^u \leq X_{\text{max}},$$

and

$$0 \leq y_t^u \leq Y_{\text{max}},$$

and

$$Z_{\text{min}} \leq z_t^u \leq Z_{\text{max}}.$$  

In this work, the energy consumption for communication, such as communication circuitry and signal processing, is ignored compared with the propulsion energy. According to [6], the propulsion energy consumption in TS $t$ is expressed as

$$e_t = \left( P_s \left( 1 + 3 \frac{v_t^h}{U_t} \right)^2 + P_m \left( \sqrt{1 + \left( \frac{v_t^h}{V_h} \right)^4 - \frac{1}{2} \left( \frac{v_t^h}{V_h} \right)^2} \right)^2 \right) t_d + \frac{1}{2} d_0 \rho_a G(v_t^h)^3 + P_k v_t^h,$$

where $P_s$, $P_m$ and $P_k$ are fixed constants and can be obtained from [6]; $U_t$ is the tip speed of the rotor blade; $V_h$ denotes the mean rotor induced velocity when hovering; $d_0$ is the main body drag ratio; $\rho_a$ is the air density; $z$ means the rotor solidity; $G$ is known as the rotor disc area; $v_t^h = \sqrt{\left(a_t^x\right)^2 + \left(a_t^y\right)^2}$ and $v_t^h = \frac{|a_t^z|}{t_d}$.

### 3.2 Channel Model

Denote the coordinate of IRS $k$ as $[x_k, y_k, z_k]$, the coordinate of UE as $[x_t^u, y_t^u]$. In this paper, the location of UE varies with time, as UE moves. Thus, the distance between UAV and IRS $k$ in TS $t$ is

$$d_{k,t} = \sqrt{(x_t^u - x_k)^2 + (y_t^u - y_k)^2 + (z_t^u - z_k)^2}.$$  

Similarly, the distance between IRS $k$ and UE in TS $t$ is given by

$$d_{k,t}^{\text{UE}} = \sqrt{(x_t^u - x_k)^2 + (y_t^u - y_k)^2 + (z_t^u - z_k)^2}.$$  

Then, for the 3-D channel model, the path loss of UAV-IRS $k$ link in TS $t$ can be denoted by $P_{k,t}^{\text{UI}}$ [3]

$$P_{k,t}^{\text{UI}} = \frac{A}{1 + \exp(-w(\theta_{k,t} - u))} + 20 \log_{10}(d_{k,t}^{\text{UI}}) + C,$$

where $A = \eta_{\text{LoS}} - \eta_{\text{NLoS}}$, $C = 20 \log_{10}(\frac{\lambda}{4\pi}) + \eta_{\text{NLoS}}$. Note that $\eta_{\text{LoS}}$ and $\eta_{\text{NLoS}}$ are variables related to the LoS and NLoS links, respectively. $\theta_{k,t} = \arctan(\frac{z_{t}^u - z_k}{\sqrt{\left(x_t^u - x_k\right)^2 + \left(y_t^u - y_k\right)^2}})$ denotes the elevation angle between UAV and IRS $k$ in TS $t$. $f$, $c$ are the carrier frequency and speed of light, respectively. $u$ and $w$ are constant values determined by the environment. Thus, motivated by [3], the channel gain of UAV-IRS $k$ link in TS $t$ is denoted by $h_{k,t}^{\text{UI}} \in \mathbb{C}^{M_r M_t \times 1}$

$$h_{k,t}^{\text{UI}} = \bar{C}(d_{k,t}^{\text{UI}})^{-\frac{4}{10}} e^{-\frac{4}{10} w(\theta_{k,t} - u)} h_{k,t}^{\text{UI}},$$

where $\bar{C} = 10^{-\frac{6}{10}}$, $\bar{A} = A \left(\frac{15}{10}\right)$. $h_{k,t}^{\text{UI}} \in \mathbb{C}^{M_r M_t \times 1}$ is the LoS component [39], [40]

$$h_{k,t}^{\text{UI}} = \left[1, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}}, \ldots, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}} \right]^T \otimes \left[1, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}}, \ldots, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}} \right]^T,$$

in which $\lambda$ is the carrier wavelength, $\delta$ is the antennas separation distance. $\phi_{k,t} = \frac{x_t^u - x_k}{\delta_k M_t}$, $\psi_{k,t} = \frac{y_t^u - y_k}{\delta_k M_t}$ represent the cosine, sine values of the horizontal, vertical angles of arrival (AoA) of the signal from the UAV to IRS $k$ in TS $t$, respectively.

Furthermore, the channel gain of IRS $k$ - UE link in TS $t$, is denoted by $h_{k,t}^{\text{IE}} \in \mathbb{C}^{M_r M_t \times 1}$

$$h_{k,t}^{\text{IE}} = \sqrt{\frac{\mu}{(d_{k,t}^{\text{IE}})^{\alpha_{\text{IE}}}}} h_{k,t}^{\text{IE}},$$

in which $\mu$ is the path loss at the reference distance $1 m$, $\alpha_{\text{IE}}$ is the path loss exponent. $h_{k,t}^{\text{IE}} \in \mathbb{C}^{M_r M_t \times 1}$ is the LoS component which is given by [40]

$$h_{k,t}^{\text{IE}} = \left[1, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}}, \ldots, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}} \right]^T \otimes \left[1, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}}, \ldots, e^{-\frac{\phi_{k,t} \psi_{k,t}}{\delta_k M_t}} \right]^T,$$

where $\phi_{k,t} = \frac{x_t^u - x_k}{\delta_k M_t}$, $\psi_{k,t} = \frac{y_t^u - y_k}{\delta_k M_t}$ represent the cosine, sine values of the horizontal, vertical angles of departure (AoD) of the signal from IRS $k$ to UE in TS $t$, respectively. Similar to [40], we denote each of IRSs has $M_r \times M_t$ reflecting elements, each of which can passively adjust its phase shift $\theta_{k,m,t} \in [-\pi, \pi]$. Thus, the diagonal phase shift matrix of IRS $k$ in TS $t$ can be expressed as $\Theta_{k,t} = \text{diag}(e^{\theta_{k,1,1,t}}, \ldots, e^{\theta_{k,m,t}}) \in \mathbb{C}^{M_r M_t \times M_r M_t}$. 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Then, the achievable data rate of UAV - IRS $k$ - UE link in TS $t$ is

$$R_{k,t} = B \log_2 \left( 1 + \frac{P(h_{k,t}^{IE})^T (\theta_{k,t}) h_{k,t}^{UL}}{B \sigma^2} \right).$$

(12)

where $P$ and $\sigma^2$ are the transmission and noise power respectively. $B$ is the bandwidth.

In this paper, assume that the UE is served with a time-division-multiple-access (TDMA) mode, where only one IRS is selected in each TS. This is very useful to save the energy consumption of the IRSs, as only one IRS may switch off at each time, whereas other IRSs may be switched off or in the sleep mode. We denote $c_{k,t} = \{0, 1\}$ as the schedule factor between UE and IRS $k$ in TS $t$, where $c_{k,t} = 1$ means IRS $k$ is selected by the UE and otherwise $c_{k,t} = 0$. Then, the schedule scheme is described as follows:

$$c_{k,t} = \begin{cases} 1, & k = \arg \min (d_{k,t}^{IE}) \text{,} \\ 0, & \text{otherwise}, \end{cases}$$

(13)

where $d_{k,t}^{IE} = \{d_{k,t}^{IE}, \forall k \in K\}$ denotes the set of distances between UE and IRSs in TS $t$. Additionally, one may have

$$\sum_{k=1}^{K} c_{k,t} = 1, \forall t \in T.$$

(14)

which means that only one IRS is selected at each time. Note that other selection schemes may also be applied. For example, one may select IRS based on the cascaded channel between UAV and UE. If the selection scheme is determined, i.e., (13), then (14) may not be needed.

### 3.3 Problem Formulation

In this paper, we aim to maximize the energy efficiency of UAV, including the data rate of UE and energy consumption of UAV, which can be formulated as the following optimization problem

$$\mathcal{P}: \max_{\Theta} \sum_{t=1}^{T} \sum_{k=1}^{K} c_{k,t} R_{k,t} / e_t,$$

subject to:

$$-x_{\min} \leq x_t \leq x_{\max},$$

(15b)

$$-y_{\min} \leq y_t \leq y_{\max},$$

(15c)

$$u_{\min} \leq u_t \leq u_{\max},$$

(15d)

$$-x_t^{u} \leq X_{\max},$$

(15e)

$$0 \leq y_t^{u} \leq Y_{\max},$$

(15f)

$$Z_{\min} \leq z_t \leq Z_{\max},$$

(15g)

$$-\pi \leq \theta_{k,m,c,t} < \pi,$$

(15h)

$$c_{k,t} = \begin{cases} 1, & k = \arg \min (d_{k,t}^{IE}) \text{,} \\ 0, & \text{otherwise}. \end{cases}$$

(15i)

where $\Theta = \{\theta_{k,t}, \forall k \in K, t \in T\}$ and $Z = \{a_t^u, a_t^v, a_t^\theta\}, \forall t \in T$. It is quite difficult to solve the above problem in general since it is non-convex. Thus, we first propose a DQN-based algorithm to tackle the trajectory of UAV by discretizing the variables $Z$. This has advantages in terms of training time, although it may result in a little bit of performance loss. We also propose a DDPG-based algorithm to optimize $Z$ with continuous actions for better performance. We also show a low-complexity phase alignment scheme to optimize $\Theta$.

### 4 PROPOSED ALGORITHMS

#### 4.1 DQN-Based Algorithm for Discrete Cases

In this subsection, we show the DQN-based algorithm. We first introduce the state, action and reward. Then, we model the whole IRS-aided UAV communication system as an environment. It is assumed that the agent is employed for interacting with the environment for the purpose of finding the optimal actions that can maximize the accumulated rewards $R_t = \sum_{t'=t}^{T} \gamma^{t-t'} r_{t'}$ within a sequence of states, where $\gamma \in [0, 1]$ is the discount factor. We define the state $s_t$, the action $a_t$, and the reward $r_t$ in TS $t$ as follows.

1. **State $s_t$:** the state of agent in TS $t$ has the following components:
   a) the coordinate of UAV: $[x_t^u, y_t^u, z_t^u]$.
   b) UAV’s remaining energy level: $e_{\max} - \sum_{t'=1}^{t} e_{t'}$, where $e_{\max}$ is the maximal energy level of UAV.
   c) the index of TS: $t$.
   d) the coordinate of UE: $[x_t^u, y_t^u]$.
   e) the set of IRS’ coordinates: $\{x_k, y_k, \forall k \in K\}$.

2. **Action $a_t$:** we define the flying distances of UAV in TS $t$ as action, which is $a_t = [a_t^u, a_t^v, a_t^\theta]$.

3. **Reward $r_t$:** we define the reward function as follows:

$$r_t = \frac{\sum_{k=1}^{K} c_{k,t} R_{k,t} - p}{e_t},$$

(16)

where $p$ is defined as the penalty if the UAV flies out of the target area, i.e., (15 e), (15 f) or (15 g) are not satisfied.

Motivated by the work that is done in [36], here we propose the DQN-based algorithm for optimizing the UAV’s trajectory, whose overall architecture is shown in Fig. 2. In DQN, there is an agent which controls the UAV for interacting with the environment. We assume there are two DNNs named the evaluation network and target network. Note that the target network has the same structure as the evaluation network but it only updates periodically. First, the agent receives the state $s_t$ from the environment and sends it to the evaluation network, which generates the Q-values.

---

**Fig. 2. Architecture of DQN algorithm.**
The Q-values of all actions. Based on the Q-values and following an $\epsilon$-greedy policy, the action $a_t$ is generated. After that, the reward $r_t$ is obtained from the environment. It is worth mentioning that the proposed DQN-based algorithm can only optimize the UAV’s trajectory in the finitely discrete action space. Hence, we define the action space in each of TS as $\mathcal{A}$, which has the following actions:

$$\mathcal{A} = \{ [x_{\text{max}}, 0, 0], [-x_{\text{max}}, 0, 0], \frac{x_{\text{max}}}{2}, 0, 0], [-\frac{x_{\text{max}}}{2}, 0, 0], [0, y_{\text{max}}, 0], [0, -y_{\text{max}}, 0], [0, \frac{y_{\text{max}}}{2}, 0], [0, -\frac{y_{\text{max}}}{2}, 0], [0, 0, z_{\text{max}}], [0, 0, -z_{\text{max}}], [0, 0, 0] \}.$$  

Then, the transition, which consists of $\{s_t, a_t, r_t, s_{t+1}\}$ is stored into an experience replay memory. When the experience replay memory has enough transitions, the learning procedure starts. A mini-batch randomly samples $M$ transitions to train the evaluation network. Precisely, given the Q-values $Q(s, a)$ from the evaluation network and the maximal Q-values $\max Q(s', a')$ from the target network, the loss function can be calculated for updating the evaluation network, which can be expressed as:

$$L_i(\delta_t) = E_{s,a} \left[ \left( r + \gamma \max_{a'} Q(s', a' | \delta_{t-1}) - Q(s, a | \delta_t) \right)^2 \right],$$  

where $\delta$ is the parameter of DNN, and $i$ is the index of iteration.

In Algorithm 1, we provide the overall pseudo code of the proposed DQN algorithm. From Line 1 to 2, we initialize the evaluation, target networks and the experience replay memory. During each episode, we first initialize the state $s_0$. Then, in each TS, the agent follows an $\epsilon$-greedy policy to generate $a_t$.

Precisely, the agent selects $a_t$ that has the maximal Q-value with probability $\epsilon$, or randomly selects $a_t$ from $\mathcal{A}$ with probability $1 - \epsilon$. The energy consumption of UAV is calculated by Eq. (4). Note that in Line 11, the selection of IRS is based on Eq. (13), and the optimization of phase shifts is introduced in Section 4.3. Then, the reward is obtained by Eq. (16). In Line 13, the transition will be stored into experience replay memory. From Line 14, the learning process starts with randomly sampling $M$ transitions from memory for training the evaluation network, whose parameter is updated by Eq. (18). Finally, the target network is also updated periodically.

Algorithm 1. DQN-Based Algorithm

1. Initialize evaluation, target networks with parameters $\delta$;
2. Initialize experience replay memory;
3. for Episode = 1, 2, ..., $N_{\text{ep}}$ do
4. Initialize state $s_0$;
5. for TS $t = 1, 2, ..., T$ do
6. Obtain $s_t$;
7. Select $a_t = \arg\max Q(s_t, a)$ with probability $\epsilon$;  
8. Randomly select $a_t$ from $\mathcal{A}$ with probability $1 - \epsilon$;
9. Execute $a_t$;
10. Calculate the energy consumption of UAV $e_t$ from Eq. (4);
11. Obtain the optimized phase shifts of selected IRS $k$ according to Section 4.3;
12. Calculate $r_t$ according to Eq. (16);
13. Store transition $\{s_t, a_t, r_t, s_{t+1}\}$ into experience replay memory;
14. if the learning process starts then
15. Randomly sample $M$ transitions from experience replay memory;
16. Update evaluation network from Eq. (18);
17. Update target network periodically;
18. end if
19. end for
20. end for

4.2 DDPG-Based Algorithm for Continuous Cases

In this subsection, we show the DDPG-based algorithm for tackling the continuous case and optimizing the UAV’s trajectory, which applies the well-known actor-critic approach. We also show the architecture of DDPG algorithm in Fig. 3. There are two DNNs named actor network with function $a = \pi(s | \delta^a)$ and critic network with function $Q(s, a | \delta^Q)$ respectively. Note that $\pi(\cdot)$ maps the state and action, $Q(\cdot)$ is the approximator for generating Q-value with the given the state-action pairs. Also, there are two target networks with function $\pi'(\cdot)$ and $Q'(\cdot)$, which have the same structure with actor and critic networks, respectively. The agent receives the state $s_t$ from the environment and sends the action $a_t$ generated by its actor network. Then, the transition is stored into the experience replay memory. When the learning process starts, $M$ transitions are sampled to train the actor and critic networks. Precisely, given the states $s$ and actions $a$, the critic network generates the Q-values $Q(s, a)$ for calculating the policy gradient [37], which is expressed as

$$\nabla_{\delta^Q} J = E \left[ \nabla_a Q(s, a | \delta^Q) \right]_{s = s_t, a = \pi(\cdot | \delta^Q)} \cdot \nabla_{\delta^Q} \pi(s | \delta^Q) \big|_{s = s_t}.$$
Once the policy gradient is calculated, the parameter of actor network is enabled to be updated. Furthermore, the critic network is trained by the loss function [37] as

\[
L(\delta^q) = \frac{1}{M} \sum_{m=1}^{M} \left( y_m - Q(s_m, \pi(s_m|\delta^q)|\delta^q)^2 \right),
\]

(20)

where \( m \) is the index of transitions in mini-batch, and \( y_m = r_m + \gamma Q(s'_m, \pi'(s'_m|\delta^q)|\delta^q). \)

Algorithm 2. DDPG-Based Algorithm

1. Initialize actor \( \pi(\cdot) \) and critic \( Q(\cdot) \) network with parameters \( \delta^\pi \) and \( \delta^Q \) respectively;
2. Initialize target networks \( \pi'(\cdot), Q'(\cdot) \) with parameters \( \delta'^\pi = \delta ^\pi, \delta'^Q = \delta^Q \);
3. Initialize experience replay memory;
4. for Episode = 1,2,...,Nep do
5. Initialize state \( s_t \);
6. for TS = 1,2,...,T do
7. Obtain \( s_t \);
8. Select \( a_t = \pi(s_t|\delta^\pi) + \omega N' \);
9. Execute \( a_t \);
10. Calculate the energy consumption of UAV \( e_t \) from Eq. (4);
11. Obtain the optimized phase shifts of selected IRS \( k \) according to Section 4.3;
12. Calculated \( r_t \) according to Eq. (16);
13. Store transition \( \{s_t, a_t, r_t, s_{t+1}\} \) into experience replay memory;
14. if the learning process starts then
15. Randomly sample \( M \) transitions from experience replay memory;
16. Update critic network according to Eq. (20);
17. Update actor network according to Eq. (19);
18. Update two target networks with rate of \( \tau \);
19. end if
20. end for
21. end for

We further provide the pseudo code of the proposed algorithm in Algorithm 2. From Line 1 to 3, we first initialize actor and critic networks with parameters \( \delta^\pi \) and \( \delta^Q \) respectively. Besides, two target networks and the experience replay memory are initialized as well. During each training episode, the state \( s_t \) is initialized in the first TS. Then, the agent obtains \( s_t \) from environment and receives \( a_t \) generated by the actor network. Note that in Line 8, a random action noise \( N' \) is deployed and it decays with rate of \( \omega \) for better exploration. In this paper, as the activation function of output layer of actor network is \( tanh(\cdot) \), the action \( a_t \) can be expressed by \( a_t = [x_t^1, x_t^2, x_t^3, x_t^4, x_t^5] \), where \( x_t^1, x_t^2, x_t^3, x_t^4, x_t^5 \) are the output values of actor network. After executing the action \( a_t \), the consumed energy \( e_t \) of UAV is obtained from Eq. (4). In Line 11, the optimized phase shifts of selected IRS and reward \( r_t \) are obtained by Section 4.3 and Eq. (16) respectively. After that, the transition \( \{s_t, a_t, r_t, s_{t+1}\} \) is stored into the experience replay memory. When the learning process starts, the mini-batch randomly samples \( M \) transitions to train the actor and critic network by Eq. (19) and Eq. (20) respectively. Additionally, two target networks are updated with the rate of \( \tau = 0.001 \).

### 4.3 Phase Shift Optimization

Here, we show a low-complexity algorithm for optimizing the phase shifts of selected IRS. Specifically, by given the coordinates of UAV and UE in TS \( t \), the phase shift of reflecting element \( \theta_{k,m,c,t} \) of selected IRS \( k \) in TS \( t \) requires to be aligned, for maximizing the data rate of UE. According to [40], the optimal phase shift \( \theta_{k,m,c,t} \) can be calculated by

\[
\theta_{k,m,c,t} = \frac{2\pi}{\lambda} \left(d(m_t - 1)\psi_{k,t}^{IE,\psi_{k,t}^{IE}} + d(m_e - 1)\psi_{k,t}^{UI,\psi_{k,t}^{UI}}\right) \tag{21}
\]

### 5 Simulation Result

In this section, extensive simulations are conducted to evaluate the performance of the proposed algorithms. The simulation is executed in Python 3.7 and Tensorflow 1.15.0. For DQN-based algorithm, we deploy two fully-connected hidden layers with [256, 256] neurons and the AdamOptimizer is used to update the evaluation network with the rate of 0.001. While the target network is updated with 300 iterations. For DDPG-based algorithm, we also deploy two fully-connected hidden layers with [256, 256] neurons in both actor and critic networks. The AdamOptimizer is used to train the actor and critic networks with the rate of 0.001. The size of experience replay memory and mini-batch are 100000 and 128 respectively. The coordinates of IRSs are set as \([100, -100, 50], [300,400,50], [500, -100, 50]\). In each training episode, the UAV always starts to serve UE from the initial coordinate \([0, 300, 50]\). The number of TSs is set as 100. The UE (which can be an autonomous vehicle) moves with a fixed speed 6 m/s starting from the initial coordinate \([0, 150]\) to the final coordinate \([600, 150]\). Other parameters can be found in Table 2.

For comparison, we present two benchmark algorithms as follows:

- Random movement and random phase shifts (RR): In this setting, the UAV randomly selects the flying
action in each TS. Also, it randomly selects the phase shift for each reflecting element.

- Fixed movement and fixed phase shifts (FF): In this setting, the UAV moves from the initial coordinate \( \frac{1}{2}, 0, 300, 50 \) to the final coordinate \( \frac{1}{2}, 600, 0, 50 \). Additionally, the phase shift of each reflecting element is fixed as \( \frac{\pi}{2} \).

First, we depict the average reward of the proposed DQN-based and DDPG-based algorithms of the training procedure with different number of reflecting elements in Fig. 4, where the number of IRSs is set to 3 and the number of the reflecting elements is the same for the IRSs. As shown in Fig. 4a, one can see for different number of reflection elements, the training curves of average rewards always remain negative at the beginning. This is because the UAV may have poor performance, such as flying out of the target area, resulting in negative reward. After that, as the networks start to converge, the average rewards increase and eventually remain stable, which indicate that the system find the best performance. Besides, one can observe that as the number of reflecting elements increases, the average rewards increase as well. Then, in Fig. 4b, we depict the average rewards of the proposed DDPG-based algorithm versus the number of training episodes, which have the similar trend as DQN-based solution in Fig. 4a. It is worth noting that when the numbers of reflecting elements are the same, DDPG-based solution achieves higher reward than DQN-based solution, as expected. This is because for DQN-based algorithm, it only tries limited set of actions, whereas DDPG-based solutions optimize the variables continuously.

When the training is done, the networks in DQN and DDPG are saved for testing. Here, we also give the complexity of proposed DQN and DDPG-based algorithms in testing phase. Specifically, as the fully-connected layers are applied in the experiments, the complexity for networks in DQN and DDPG is \( \mathcal{O}(\sum_{l=1}^{L} n_{-1} n_l) \), where \( L \) denotes the number of layers and \( n_l \) is the number of neurons in \( l \)th layer. Besides, the complexity for phase shift optimization in each TS is \( \mathcal{O}(M_t M_e) \). Thus, the overall complexity for DQN and DDPG is \( \mathcal{O}(T(M_t M_e + \sum_{l=1}^{L} n_{-1} n_l)) \).

Then, we evaluate the performance of proposed DQN and DDPG-based algorithms. In Fig. 5, we depict the average energy efficiency of UAV obtained by DQN, DDPG, FF, RR respectively in one episode. Specifically, the energy efficiency of UAV obtained by DDPG consistently increases from 52 bps/J to 70 bps/J. Additionally, it is observed that for different number of reflecting elements, DDPG always achieves higher energy efficiency comparing with other algorithms. DQN performs slightly worse than DDPG, which also outperforms FF and RR.

Then, we show the 3D and 2D trajectories obtained by DQN and DDPG with different number of reflecting elements in Fig. 6. Note that in Fig. 6, dot represents UE, and triangle represents IRS. In Fig. 6a, it is observed that the UAV controlled by DQN starts to serve UE from the initial coordinate and finally flies to the appropriate altitude for achieving better performance. In Fig. 6b, one can see that as the location of UE moves, the UAV flies towards to the selected IRS and remains close to it with appropriate flying actions. Also, as shown in Figs. 6c and 6d, the UAV’s trajectory obtained by DDPG is better than the trajectory achieved by DQN, as it always tries continuous actions.

Finally, we show the training time of DQN and DDPG-based algorithms versus the number of reflecting elements.
of IRS in Fig. 7. Note that the training time will vary with different hardware platform. As shown in Fig. 7, one can see that as the number of reflecting elements increases, the training time of DQN and DDPG increases as well. Besides, DQN consistently outperforms DDPG in terms of training time, for its simpler structure.

6 CONCLUSION

In this paper, we have studied the joint optimization of UAV’s trajectory and passive phase shifts of reflection elements in the IRS-aided UAV communication system, with the consideration of the movement of UE and the selection of IRS. Our aim is to maximize the energy efficiency of the system, including the data rate of UE and the energy consumption of UAV. We have first proposed a DQN-based algorithm by discretizing the trajectory, which has advantage in terms of training time but has performance loss, which may be suitable for the cases that is sensitive to the training time. Then, for achieving the better performance, we have further applied a DDPG-based algorithm, which can optimize the system’s variables continuously. The experimental results have proved that the proposed algorithms achieve better performance then other traditional solutions.
REFERENCES

[1] F. Jiang, K. Wang, L. Dong, C. Pan, W. Xu, and K. Yang, “AI driven heterogeneous MEC system with UAV assistance for dynamic environment: Challenges and solutions,” IEEE Netw., vol. 33, no. 4, pp. 408–412, Jul./Aug. 2019.

[2] B. Yang, X. Cao, C. Yuen, and L. Qian, “Offloading optimization in edge computing for deep learning enabled target tracking by internet-of-UVAs,” IEEE Internet Things J., vol. 8, no. 12, pp. 9878–9893, Jun. 2020.

[3] A. Al-Hourani, S. Kandeepan, and S. Lardner, “Optimal LAP altitude for maximum coverage,” IEEE Wireless Commun. Lett., vol. 3, no. 6, pp. 569–572, Dec. 2014.

[4] Q. Wu and R. Zhang, “Common throughput maximization in UAV-enabled OFDMA systems with delay consideration,” IEEE Trans. Commun., vol. 66, no. 12, pp. 6614–6627, Dec. 2018.

[5] H. Ren, C. Pan, K. Wang, W. Xu, M. Elkashlan, and A. Nallanathan, “Joint transmit power and placement optimization for URLLC-enabled UAV relay systems,” IEEE Trans. Veh. Technol., vol. 69, no. 7, pp. 8003–8007, Jul. 2020.

[6] Y. Zeng, J. Xu, and C. Pan, “Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach,” IEEE J. Sel. Areas Commun., vol. 36, no. 9, pp. 2059–2070, Sep. 2018.

[7] X. Yu, L. Xiao, C. Wang, and H. Dai, “Distributed cellular communications with deep reinforcement learning against jamming,” IEEE Wireless Commun., vol. 27, no. 4, pp. 48–53, Aug. 2020.

[8] Z. Yang, C. Pan, K. Wang, and M. Shikh-Bahaei, “Energy efficient resource allocation in UAV-enabled mobile edge computing networks,” IEEE Trans. Wireless Commun., vol. 18, no. 9, pp. 4576–4589, Sep. 2019.

[9] L. Wang, K. Wang, C. Pan, W. Xu, N. Aslam, and L. Hanzo, “Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing,” IEEE Trans. Cogn. Commun. Netw., vol. 7, no. 1, pp. 73–84, Mar. 2021.

[10] L. Wang, K. Wang, C. Pan, W. Xu, N. Aslam, and A. Nallanathan, “Deep reinforcement learning based dynamic trajectory control for UAV-assisted mobile edge computing,” IEEE Trans. Mobile Comput., early access, Feb. 16, 2021, doi: 10.1109/TMC.2021.3089697.

[11] W. Huang et al., “Joint power, altitude, location and bandwidth optimization for UAV with underlaid D2D communications,” IEEE Wireless Commun. Lett., vol. 8, no. 2, pp. 524–527, 2019.

[12] C. Zhan, Y. Zeng, and R. Zhang, “Energy-efficient data collection in UAV enabled wireless sensor network,” IEEE Wireless Commun. Lett., vol. 7, no. 7, pp. 328–331, Jul. 2018.

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

[14] C. Huang, C. Pan, H. Ren, K. Wang, and A. Nallanathan, “Intelligent reflecting surface aided multigroup multicast MISO communication systems,” IEEE Trans. Signal Process., vol. 68, pp. 3236–3251, 2020.

[15] C. Pan et al., “Multicell MIMO communications relying on intelligent reflecting surfaces,” IEEE Trans. Wireless Commun., vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[16] C. Pan et al., “Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[17] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, “Latency minimization for intelligent reflecting surface aided mobile edge computing,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1839–1850, Aug. 2020.

[18] S. Abeywickrama, R. Zhang, and C. Yuen, “Intelligent reflecting surface: Practical phase shift model and beamforming optimization,” in Proc. IEEE Int. Conf. on Commun., 2020, pp. 1–6.

[19] A. Zappone, M. Di Renzo, F. Shams, X. Qian, and M. Debbah, “Overhead-aware design of reconfigurable intelligent surfaces in smart radio environments,” 2020, arXiv:2003.02538.

[20] C. Huang, R. Mo, and C. Yuen, “Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1839–1850, Aug. 2020.

[21] S. Li, B. Duo, X. Yuan, Y.-C. Liang, and M. Di Renzo, “Reconfigurable intelligent surface assisted UAV communication: Joint trajectory design and passive beamforming,” IEEE Wireless Commun. Lett., vol. 9, no. 5, pp. 716–720, May 2020.

[22] M. Da, M. Ding, and M. Hassan, “Enhancing cellular communications for UAVs via intelligent reflective surface,” 2019, arXiv:1911.07631.

[23] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, Cambridge, MA, USA: A Bradford Book, 2018.

[24] V. Mnih et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, 2015.

[25] T. P. Lillicrap et al., “Continuous control with deep reinforcement learning,” 2015, arXiv:1502.05477.

[26] A. G. Barto, R. S. Sutton, and C. W. Anderson, “Neurodynamic adaptive elements that can solve difficult learning control problems,” IEEE Trans. Syst., Man, Cybern., vol. SMC-13, no. 5, pp. 529–533, Sep./Oct. 1983.

[27] H. Ren, C. Pan, K. Wang, Y. Deng, M. Elkashlan, and A. Nallanathan, “Achievable data rate for URLLC-enabled UAV systems with 3-D channel model,” IEEE Wireless Commun. Lett., vol. 8, no. 6, pp. 1587–1590, Dec. 2019.

[28] Z. Wei et al., “Sum-rate maximization for IRS-assisted UAV OFDMA communication systems,” IEEE Trans. Wireless Commun., vol. 20, no. 4, pp. 2530–2550, Apr. 2020.

Liang Wang received the BEng degree, in 2014, the MSc degree, in 2015, and the PhD degree from Northumbria University, U.K., in 2022. He is currently working as a research fellow with Cranfield University, U.K. His research interests include UAV communication, mobile edge computing, physical layer security and machine learning.
Kezhi Wang received the PhD degree in engineering from the University of Warwick, U.K., in 2015. He was a senior research officer in University of Essex, U.K. from 2015-2017. Currently he is with Department of Computer and Information Sciences, Northumbria University, U.K. His research interests include mobile edge computing, intelligent reflection surface, and machine learning.

Cunhua Pan received the BS and PhD degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he was a research associate with the University of Kent, U.K. He held a post-doctoral position with the Queen Mary University of London, U.K., from 2016 and 2019. From 2019 to 2021, he was a lecturer with the same university. From 2021, he is a full professor in Southeast University. His research interests mainly include reconfigurable intelligent surfaces (RIS), intelligent reflection surface (IRS), ultra-reliable low latency communication (URLLC), machine learning, UAV, Internet of Things, and mobile edge computing. He has published more than 120 IEEE journal papers. He is currently an editor of IEEE Wireless Communication Letters, IEEE Communications Letters and IEEE Access. He serves as the guest editor for IEEE Journal on Selected Areas in Communications on the special issue on xURLLC in 6G: Next Generation Ultra-Reliable and Low-Latency Communications. He also serves as a leading guest editor of IEEE Journal of Selected Topics in Signal Processing Special Issue on Advanced Signal Processing for Reconfigurable Intelligent Surface-aided 6G Networks, leading guest editor of IEEE Vehicular Technology Magazine on the special issue on Backscatter and Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G, leading guest editor of IEEE Open Journal of Vehicular Technology on the special issue of Reconfigurable Intelligent Surface Aided Communications for 6G and Beyond. He is a workshop organizer in IEEE ICC 2021 on the topic of Reconfigurable Intelligent Surfaces for Next Generation Wireless Communications (RIS for 6G Networks), and workshop organizer in IEEE Globecom 2021 on the topic of Reconfigurable Intelligent Surfaces for future wireless communications. He is currently the workshops and symposia officer for Reconfigurable Intelligent Surfaces Emerging Technology Initiative. He is workshop chair for IEEE WCNC 2024, and TPC co-chair for IEEE ICT 2022. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the student travel grant chair for ICC 2019. He received the IEEE ComSoc Leonard G. Abraham Prize, in 2022.

Nauman Aslam received the PhD degree in engineering mathematics from Dalhousie University, Canada, in 2008. He is a professor with the Department of Computer and Information Science, Northumbria University, UK. Before joining Northumbria University as a senior lecturer, in 2011, he worked as an assistant professor with Dalhousie University, Canada. He is leading the Cyber Security and Network Systems (Cyber-Nets) research group, Northumbria University. His research interests cover diverse but interconnected areas related to communication networks. His current research efforts are focused at addressing problems related to wireless body area networks and IoT, network security, charging management of electric vehicles and application of Artificial Intelligence (AI) in communication networks. He has held, as PI and CoI, grants totalling over £5 million from funders such as EU, EPSRC, Innovate UK, as well as industry, and is the co-author of more than 150 peer-reviewed publications. He is the chair of IEEE UK and Ireland Communications Society (ComSoc) chapter.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.