IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 August 2023, accepted 9 September 2023, date of publication 14 September 2023,
date of current version 21 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3315660

RESEARCH ARTICLE

Capacity Enhancement of Flying-IRS Assisted 6G THz Network Using Deep Reinforcement Learning

SHEREEN S. OMAR, AHMED M. ABD EL-HALEEM, (Member, IEEE), IBRAHIM I. IBRAHIM, AND AMANY M. SALEH
Electronic and Communication Engineering Department, Helwan University, Helwan, Cairo 11792, Egypt

Corresponding authors: Shereen S. Omar (shereenomaar10@gmail.com) and Amany M. Saleh (amany_mohamed_saleh@h-eng.helwan.edu.eg)

ABSTRACT Terahertz communication networks and Intelligent Reflecting Surfaces (IRS) exhibit significant potential in advancing Sixth-Generation (6G) wireless networks, these technologies enable support of ultra-data transmission and exploding network capacity. Motivated by the above facts, this paper considers the flying IRS assisted Unmanned Aerial Vehicles (UAV) in THz communication network. To that aim, we proposed algorithm named a (Fly-IRS) aided-THz communication network by jointly optimizing the optimal user Grouping and the IRS phase shifting, optimal UAV's location optimization are being explored to achieve the system data rate maximization, enhancing the system capacity, and minimizing the Outage Probability (OP) to provide a better satisfied user ratio. The formulated problem is decomposed into two subproblems, an iterative algorithm based on modified K-means clustering algorithm is proposed to solve the first sub-problem: the optimizing user Grouping, while Deep Deterministic Policy Gradient (DDPG) optimizes the IRS phase shift and optimal UAV's location optimization. Finally, simulation results demonstrate that the proposed algorithm can maximize the system's data rate by up to 95% and improves the capacity of the system on average by 94% compared to benchmark algorithms.

INDEX TERMS Sixth-generation (6G) network, terahertz (THz) communication, unmanned aerial vehicles (UAVs), intelligent reconfigurable surface (IRS), user grouping design, IRS phase shift and UAV location design, deep deterministic policy gradient (DDPG).

I. INTRODUCTION

Because of the necessity for high-speed data links whenever and wherever they are required, wireless communication systems have evolved significantly over the last many decades [1], [2], [3]. For the future sixth generation (6G) systems scaled up beyond wide-band MIMO technology. Ultra-large bandwidth transmission, where large-scale antenna arrays are used to transmit high frequency signals with ultra-large bandwidths and design of the hybrid transmit precoder (TPC) scheme, called spatial-wideband effect causes grave beam squint, [4], [5]. Furthermore, 6G application like wireless data centers, holographic tracking systems, and on-chip communications require networks to be able to provide ubiquitous and diverse services while also satisfying rising bandwidth requirements, researchers and technology developers began looking into the only available gap in the frequency spectrum [6], [7], [8]. To satisfy the growing need for greater rates of data and novel spectral bands, the Terahertz (THz) (0.1–10 THz) communication band that supports applications and THz links could ensure low-latency and dependable exchange of data by leveraging its extremely high rate of data with a high quality of service when properly deployed. THz wireless is considered one of the most promising techniques for enabling ultra-high-speed communication, which theoretically achieves up to some THz, resulting in a possible capacity in the order of terabits per second [9]. Shorter distances are supported among the transmitter and receiver at THz because of the significant path loss. But this also correlates to THz's reduced power requirements and perhaps improved efficiency of energy, all of which contribute to the THz communication features [10]. THz network is also well suited to support emerging applications such as high- quality streaming of videos, augmented

The associate editor coordinating the review of this manuscript and approving it for publication was Ding Xu.

and virtual reality, and chip-based wireless networks. Additionally, owing to their ability to support non-line-of-sight propagation, the THz signals have proven their suitability for downlink communications [11], [12]. Effective wireless network integration might be achieved, in particular, by utilizing THz-capable flying platforms in the communication network, like Unmanned Aerial Vehicles (UAVs), to provide seamless coverage and high bandwidth to ground users. UAVs have become in great demand on wireless networks in areas with no or inadequate land infrastructure due to their flexibility and mobility [13], [14]. For the important issues of UAVs is the management and the obstacle recognition, some studies have proposed various approaches such as the rapidly expansion strategy of randomized trees and sample-based approaches [15], [16]. In order to increase coverage and improving the reliability and data rate of wireless networks, UAVs essentially serve as aerial base stations [17], [18]. More specifically, THz communications are expected to contribute to the 6G wireless systems massive bandwidth requirement for high rates of data. However, THz communications performance is significantly hampered by severe path attenuation, transceiver antenna misalignment, and imperfect hardware. However, the utilized of UAVs in ultra-data transmission networks remains limited and still faces many challenges. The most important challenge is that usually have limited UAVs battery lifetime usually due to the aircraft weight and size constraint. In particular, in the crowded environment the coverage and connectivity difficulties remain the problem for UAV communications. To overcome these challenges and compensate for the abovementioned weakness of UAVs, they utilized the Intelligent Reconfigurable Surface (IRS) for THz communications systems. To overcome this challenge and compensate for the above-mentioned weakness of UAVs, they utilized the IRS for THz communications. In particular, the UAV can be equipped with passive IRS, which is one of the savviest, easiest, and most energy-efficient ways. Recent developments in wireless communications have seen the introduction and implementation of IRS's new technologies to improve the performance of communication [19], [20]. The IRS's most important advantage is that it consumes far less power compared to traditional processing of signals for a BS. This is going to improve the communications between UAV-based IRS systems energy effectiveness. To this end, IRS phase shift designs have been proposed in [21] has studied the reflection design and channel estimation problem for IRS aided mmWave communications to maximize information rate in the presence of the beam-squint effect. Reference [22] IRS can connect to an airborne UAV in the setting of 6G wireless communication THz networks to actualize an effective method for assisting heterogeneous customers from the sky [23]. This interaction generates a novel variety of the IRS class known as flying IRS, which comprises an airborne vehicle serving as a transporter that carries the IRS that functions as a reflector. Which allows for flexible IRS deployment and 360° panoramic full-angle reflection, which can improve the area coverage extension in THz communication system, also, the cost of deploying flying IRS in the system is lower than the cost of deploying more BSs [24], [25]. Despite the recent application and deployment of IRS in THz communications, they are still in their infancy [26], [27]. For example, to combat short-range communications in terahertz networks, the researchers of [28] investigated the various access scenarios of the multiple IRS systems. The two goals to be accomplished by adjusting the phase shift of the multiple IRS optimization are maximizing the data received rate of a desired user while being interference with by the second user and maximizing the total rate for both users. That is a non-convex challenge and more difficult than using Deep Reinforcement Learning (DRL) to solve both goals. Recently, the rather attractive machine learning-based approach has been developed, which is based on Reinforcement Learning (RL) and the use of multilayer neural networks greatly improved the estimation accuracy. Such as the field of adaptive control technology, including Robust Integral of the Sign of the Error (RISE) control, and data compression and damage evaluation method of underground pipeline is investigated in [29] and [30]. To further improve the coverage of the links at THz for communication. The researchers in [31] suggested a hybrid beamforming strategy for the cascaded IRS-aided networks. To tackle the non-convex optimization problem, they used the DRL algorithm. To combat the loss of propagation in the THz downlink broadcast system, they investigated the IRS beamforming matrix in the joint design and the digital beamforming at the BS. Moreover, in [32], the researchers investigated the performance of the maximum sum rate achieved with individual rate constraints by jointly optimizing the IRS location, the IRS phase shift, the distribution of THz band network, and power control for users. The non-convex problem is suggested to be solved using the block coordinate searching (BCS) algorithm. Deep Reinforcement Learning (DRL) was implemented by the authors of [33] to improve the communication effectiveness and trajectories of THz-enabled UAVs when constraints are enforced by shifting THz channel circumstances for Ground user connections. By jointly maximizing the deployment of operational UAVs and ground user's association while also reducing the transmitting power of UAVs. In [34], the authors have focused on developing algorithms for maximizing maximize the system sum rate utilities UAV equipped with IRS assistance D2D-enabled in THz wireless network. A joint power allocation design and RIS phase shift optimization problem is proposed to maximize sum rate of the system under the individual QoS, power and practical discrete phase shift constraints. After that, a phase shifts optimization problem is formulated and solved using local search method for the phase shifts.

A recent study [35] investigated the IRS in THz network. In a THz-band IRS-aided integrated sensing and communications (ISAC) system with multiple users, the authors studied joint transmit beamforming and phase shifting design optimization. To maximize active and passive precoding, the

researchers combined DRL with a primal-dual proximal policy. The efficacy of DRL in solving the high-dimensional nonconvex issue to increase the sum rate of users' communication is confirmed by the simulation findings. However, [35] only takes narrowband scenarios into account without making use of the key benefit of ultra-wide bandwidths in THz.

Novelty and Contributions. In this work, we aim to study quantitatively the applicability of THz communications network in blockage-rich mobile environments and utilized the Flying IRS assists-THz communication network based Deep Reinforcement Learning method. Consequently, we considered the downlink transmission model in THz network system operating in outdoor environment where the direct link among the Base Station (BS) and Ground Users (GUs) is blocked by surrounding obstructions. Therefore, we proposed algorithm named a (Fly-IRS) aided-THz communication system by jointly optimize the optimal user Grouping and IRS phase shifting with the optimal UAV's location, to maximize the system data rate, enhance the capacity of system. On the other hand, the proposed algorithm is minimizing the Outage Probability (OP) tends to provide superior performance in the satisfied user ratio to improve the performance of the system.

- Considering flying IRS aided-THz communication systems, we proposed a cost-effective algorithm named a (Fly-IRS) algorithm aided-THz wireless communications network, the joint optimal user Grouping and IRS phase shifts with optimal UAV's location optimization algorithm is proposed, to maximize the system data rate, and enhance the system capacity.
- We decomposed the optimization problem into two sub-problems: the user Grouping sub-problem and IRS phase shifts, optimal UAV's location optimization sub-problem.
- Two optimization problem phases are considered to maximize the total data rate of THz communications system, and it is formulated as a non-convex non-linear mixed integer problem.
- The first sub-problem is the optimal user Grouping, was solved using iterative algorithm by modified K-means clustering algorithm. The DDPG algorithm is introduced into DRL framework to optimize the phase shifting of IRS and the optimal location of the UAV simultaneously, to resolve the second sub-problem.
- Simulation results are provided to demonstrate the superior performance of the proposed algorithm other compared schemes from the literature in [33], [34], and [35] on the system data rate, improves the capacity of the system and minimizing the Outage Probability to provide a better satisfied user ratio.

The rest of this paper is structured as follows: Section II. presents the system model, while Section III. describes how to formulate problems. Section IV. FRAMEWORK OF The Proposed Algorithm Ans Policy, while Section V exploits the results of the simulation, Challenges and Future Directions are provided in Section VI, Finally, Section VII. presents the conclusion.

II. SYSTEM MODEL

For a THz-enabled multi-IRS-assisted UAV wireless communication network, the system model of downlink data transmission is introduced in this section. The detailed architecture of the system model is shown in Figure (1), which consist of one BS operating at the THz frequency band, equipped with multiple antennas. Assuming that the direct links among the BS and the downlink GUs are blocked by buildings or other environmental impediments, we employ Fly-IRS proposed system downlink data transmission service for all users. The BS is located at the midpoint of an urban area denoted by A and serve GUs users are randomly distributed in that area denoted by M, which are partitioned into different groups to enhance the system's performance. The communications between the BS and the GUs have been assisted by a Flying IRSs that is considered to be a Uniform Planar Array (UPA). The overall number of reflective elements is $N = N_x N_y$, where N_x and N_y are the number of IRS reflecting elements along the X-axis and Y-axis respectively. The IRS learns the optimal method to reflect incident signals by adjusting the phase shift, which helps improve the THz network's performance. To ease future analysis, a relative coordinate Fly-IRS is denoted by $q(x, y)$. Suppose that the UAV flies at a constant altitude over an area where the duration of flight and the energy consumption of the UAV are neglected in our work.

Without loss of generality, considered the location (3D coordinates) of BS can be defined as q_B=($x^{BS} = 0, y^{BS} = 0, h^{BS}$) and the coordinates of the user location is q_M=(x^{GU}, y^{GU}, h^{GU}) Let UAV is coordinates q_I=($x^{IRS}, y^{IRS}, h^{IRS}$) respectively.

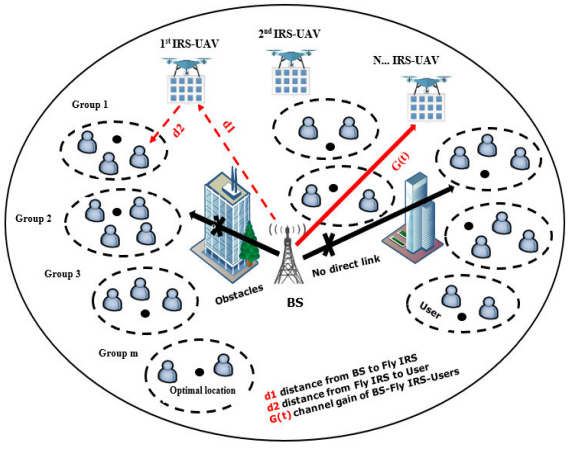

FIGURE 1. System model of Fly-IRS aided-THz communications.

A. CHANNEL MODELING

Spreading and molecule absorption losses determine signal propagation in the THz band [36]. As a result, the channel transfer function is as follows:

$$H(f, d) = \frac{c}{4\pi f d} e^{-0.5\, k(f) d} \tag{1}$$

where f is the carrier frequency, c is the speed of light, the term $e^{-0.5k(f)d}$ represents the channel path-loss due to the molecular absorption and the concentration of water vapor molecules in the air.

Detailed equation (2): The $K(f)$ denotes the absorption loss coefficient which, based on the transmission frequency represents as:

$$K(f) = \frac{P}{P_{stp}} \frac{T_{stp}}{T} \sum Q^{i,g} \sigma^{i,g}(f) \tag{2}$$

where in p_{STP} is represents the standard pressure, T_{STP} is the standard temperature, p is the pressure, and T represents the temperature. Then, $Q^{i,g}$ is indicated as the transmission environment, and $\sigma^{i,g}(f)$ represents the total number of molecules per unit volume at the frequency f [37].

Also, $d = d_1 + d_2$ represents the Euclidean distance from BS to the Fly-IRS and the distance from the (Fly-IRS to the GUs, where d_1 is the distance among the BS and the first reflector of IRS, and d_2 is the distance among the first reflector of IRS and the users, which can be calculated by:

$$d_1 = \sqrt{\left(x^{BS} - x^{IRS}\right) + \left(y^{BS} - y^{IRS}\right) + \left(h^{BS} - h^{IRS}\right)}$$
$$d_2 = \sqrt{\left(x^{IRS} - x^{UE}\right) + \left(y^{IRS} - y^{UE}\right) + \left(h^{IRS} - h^{UE}\right)}$$

B. TRANSMISSION STRATEGY

Suppose that direct links between BS and GUs are blocked owing to severe blockage or a significant distance, requiring the importance of aerial platforms like flying IRSs, which assist the THz communication system and create a reliable wireless connection for data dissemination. Consider that there is an IRS deployed at a UAV parallel to the ground (Fly-IRS), with coordinates (x-axis, y-axis). Let N_x and N_y be the number of IRS reflecting elements, the total number of reflecting elements is $= (N_x N_y)$.

The channel gain of BS- (Fly-IRS)-GUs [38]:

$$G(t) = g(t)\, \Phi e_b e_i \tag{3}$$

where Φ beamforming matrix related to IRS, defined as:

$$\Phi = diag[w_n N e^{j\phi 1}, e^{j\phi 2}, \ldots..e^{j\phi M}] \tag{4}$$

Since each element of the IRS main diagonal represents both $w_n \in [0, 1]$ amplitude reflection coefficient of the reflective element of the Fly-IRS, and j is a complex number's imaginary unit. Assumed the amplitude is set as $w_n = 1$, similar to [39], and the phase shift of the $(N_x N_y)$-th element of the IRS where the value of $\phi \in [0, 2\pi]$.

The cascaded channel gain of the BS- (Fly-IRS) -GUs link is $g(t)$ at time t [39]:

$$g(t) = \frac{c}{8\sqrt{\pi^3} f r_u(t)\, r_o(t)} e^{-j2\Pi f \frac{(r_t(t)+r_o(t))}{C}}$$
$$e^{-(-0.5k(f)(r_t(t)+r_o(t)))} \tag{5}$$

As a result, $r_o(t)$ is the transmission vector from BS to the first component of the Fly-IRS. The relative phase difference among the received signal BS and the first IRS element can be represented by:

$$\theta_{N_x N_y(t)} = \frac{2\Pi f r_o(t)\, \Delta r_{N_x N_y(t)}}{|r_o(t)|} \tag{6}$$

where $\Delta r_{N_x N_y(t)}$ is indicated as the difference vector from the IRS elements.

Mathematically, the reflected signal is multiplied by the complex reflection coefficient to produce the reflected signal of IRS. The received array vector from BS to the IRS at time t can be expressed by:

$$e_b = \left[e^{-j_{\theta 1}}, e^{-j_{\theta 2}}, \ldots..e^{-j_{\theta M}}\right] \tag{7}$$

The relative phase difference among the elements of the IRSs reflected beams towards the GUs.

Thus, the transmit array vector from Fly-IRS to the k-th GU can be represented as:

$$e_i = \left[e^{-j_{\beta 1}}, e^{-j_{\beta 2}}, \ldots..e^{-j_{\beta M}}\right] \tag{8}$$

Utilize the same frequency band for multi-IRS-UAVs at the same time. The network is experiencing interference from another UAV and GUs. As a result, the signal-to-interference plus noise ratio (SINR) from every UAV to GUs during every time t could be represented as follows:

$$SNIR = \left(1 + \frac{P_t * G(t)}{\psi + B_i \left(de^{-0.5\, k(f)d}\right) \sigma^2}\right) \tag{9}$$

where ψ the network experiences interference, $G(t)$ is the channel gain, P_t is the transmit power from the UAV to their GUs, B_i is the total THz-bandwidth, and σ^2 is the additive white Gaussian noise (AWGN).

The sum data rate of users could be obtained based on the SINR as follows:

$$R_m = B_i / m log_2 \left(1 + \frac{P_t * G(t)}{\psi + B_i \left(de^{-0.5\, k(f)d}\right) \sigma^2}\right) \tag{10}$$

III. PROBLEM FORMULATION

This section formulates the system problem to maximize the data rate of the system, subject to the constraint of the user data rate requirement. Hence, the formulated optimization problem could be decomposed into two sub-problems. The first sub-problem can be solved using iterative algorithm by modified K-means clustering algorithm, to determine the optimal user Grouping. For Second sub-problem, IRS phase shift and optimal UAV's location optimization, a Deep Deterministic Policy Gradient (DDPG) algorithm was proposed to resolve this problem to near optimal solutions.

Therefore, the data rate maximization problem could be expressed as:

$$P1: \max_{\{\theta_n, q, p, \boldsymbol{w}_n\}} \sum_{m=1}^{M} R_m \tag{11}$$
$$\text{s.t: } R_m \geq R \min, \forall t \in K \tag{11a}$$

$$\sum_{m=1}^{M} P_k \leq P_{\max} \quad (11b)$$

$$w_n = 1, \forall_n \in N \quad (11c)$$

$$0 \leq \theta_n \leq 2\pi n = 1, \ldots .N \quad (11d)$$

$$q(x, y) \in \quad (11e)$$

Constraint (11a) minimum achievable rate for all the users while guaranteeing that the quality of service (QoS) for every user is satisfied. (11b) ensures that the overall transmitted power doesn't exceed the BS maximum power. The constraint (11c) implies that the IRS reflection unit is total reflection, with the amplitude of the reflection coefficient in all the IRS elements being 1.

Whereas the constraints of (11d) indicate the characteristics of the IRS reflecting matrices, every one of them is a type of IRS-reflecting phase shift matrix that reflects all transmitted signals without any loss of power. The constraint of (11e) limits the UAV's ability to fly only within a certain feasible area.

The optimization problem in (11) is a non-convex problem that is difficult and cannot be solved directly. Thus, in the following sections, we decompose the original problem into two subproblems and find the optimal user Grouping and optimal location of UAVs with an orientation of the IRS phase shift.

IV. FRAMEWORK OF THE PROPOSED ALGORITHM AND POLICY

The original problem was divided into two subproblems in this section. The first sub-problem focuses on optimal user Grouping, while the second sub-problem optimizes the joint IRS phase shift with optimal UAV location to maximize the system's average data rate. To resolve the first sub-problem, an iterative algorithm is used. For the second sub-problem, a DDPG algorithm for the DRL model is employed until the algorithm converges and search for the optimal solutions.

The following subsections provide details of the proposed algorithms.

A. PHASE ONE OPTIMAL USER GROUPING

In this section, an enhanced scheme based on the optimal user Grouping scheme is proposed.

The principle of the Grouping scheme is the task of partitioning users into groups that are classified based on shortest distance criteria to allocate elements to groups and apply data rate constraints according to predefined requirements. It is exceedingly challenging to solve the user Grouping problem optimally in large-scale THz systems. Since an exhaustively searching for the best user Grouping solution would have excessive computing complexity, a modified k-means algorithm is proposed for user Grouping based on some vital modifications to the k-means algorithm such that it can incorporate shortest distance criteria and data rate assigned constraints for each group separately.

The K-means cluster algorithm is an iterative method that attempts to make partitions in an unsupervised dataset. It has the advantages of low complexity, easy implementation, and fast convergence. Algorithm 1 describes a proposed enhanced technique that may accomplish faster convergence based on the ideal user Grouping taking into account the initial grouping settings. The election of groups in this paper is realized by these specific procedures:

Algorithm 1 Proposed User Grouping Scheme

Input: Number of users, Distance threshold, Data rate threshold

Initialization: Randomly initialize generate location for each user and Initial assign random data rate values for each user

The scheme iteration step (update each group)

1: **for** every user $m = 1, \ldots, M$ **do**

2: Calculate Euclidean distance between the user and neighboring users.

3: Assign random data rate for each user using random distribution techniques.

4: The user belongs to the group with the closest distance and maximum data rate for every group.

5: **Repeat** step 2 *to* 4 for all users until convergence.

6: **end for**

UntilGroup is formed with minimum distance between users and achieves maximum data rate for each group. **Output:** Optimal User Groups

In the first step, at the beginning of each iteration, it can calculate the minimum Euclidean distance between each user and the neighboring users. If the Euclidean distance among them is the nearest distance, the user belongs to the group acted served by the m-th Fly-IRS.

In the second step, assigned random data rate for each user. For improving the grouping of completed forms according to the total data rate assigned to the groups, there are some conditions to adjust the group coverage. The probability that a randomly chosen user's required data rate is higher than a specific data rate threshold is used to determine the group coverage probability. This KPI, in particular, tests if the forming group created as required is completed with the users belonging to this group. At the end of each iteration, and then after resetting the locations of the users, repeat the above steps until the group does not change.

According to the use of Fly-IRS proposed in THz bands, the higher the height of the UAV, the larger the communication coverage area. So, the optimal coverage radius **r** of the design area based on the altitude of the Fly-IRS and beam angles in the THz network can be calculated as follows:

$$radius\ \mathbf{r} = h_i * tan(\theta) \quad (12)$$

where θ is the half beam angle in the THz band, with a narrower antenna beamwidth of $<10^{\circ}$ [40], [41], and h_i is the height of the UAV, respectively.

The optimal maximum coverage area of an (IRS-UAV) is shown in Figure (2).

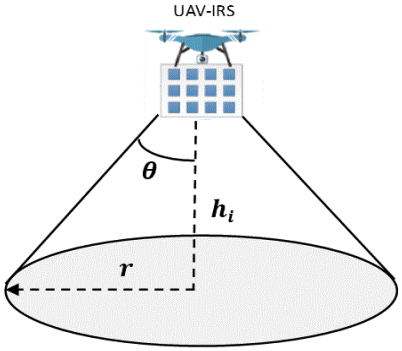

FIGURE 2. Coverage area of Fly-IRS with beam width angle θ and UAV altitude h_i in THz network.

B. PHASE TWO: JOINT OPTIMIZATION PHASE SHIFT OF IRSS AND OPTIMAL UAV LOCATION

The proposed DRL-DDPG algorithm and optimization policy framework are introduced in this subsection. Depending on the above system model, we design the joint phase shift matrices and the optimal UAV location in THz communication network. By discovering the best solutions in a dynamic environment is an efficient dynamic programming technique for solving sequential decision-making problems. In that case, with the dynamic environment, the Fly-IRS then assumes the role of the agent. The difficulty with this reinforcement learning framework arises from the fact that action and state possess continuous values. So, we proposed DRL framework based DDPG approach to select the IRS phase shift while considering the consumed beamforming training time. DDPG algorithm deal with continuous actions effectively according to parametric strategy (such as IRS passive phase shifts). Furthermore, it has neural network and Actor-Critic structure for selection continuous action [42], [43]. We try to make the DDPG choose the best IRS board using distributed training action selection policy and solving problems with continuous action at the cost of a slight performance loss, to adapt dynamic communication environment without any prior channel state information, that reduce power consume and reduce overhead training [44], [45].

The DDPG algorithm is described briefly at first, followed by specific actions, states, and rewards.

Lastly, consider how the DDPG framework may be used to resolve the formulated problem of finding the optimal location of the UAV and phase shift of IRS.

1) DDPG ALGORITHM WORKING PROCEDURE

In DDPG, the training stage begins once the experience replay buffer is filled. For the training of the four neural networks, the batch-size N_B transitions $(s_t, a_t, r_t, s_{t+1}, a_{t+1})$ are chosen as mini-batch [46], [47]. The DDPG algorithm aims to identify the optimal action that will maximize the Q-value. The Q-value function can be utilized to examine a state action pair's quality as a value function that maximizes the expected cumulative reward under an optimal policy π. It enables evaluation of the agent's actions and transitions between states. Given the state s_t, action a_t and reward r_t.

The DDPG algorithm also adds randomly generated noise N to help the agent achieve better exploration, where N is the Gaussian noise, whose dimension same that of the output action.

Also, we define the two identical networks as target Q-value $Q(\theta^{\text{train}}|s_t, a_t)$ and training Q-value $Q(\theta^{\text{target}}|s_t, a_t)$ and the loss function is used to define the differences between them. Suppose that the training and the target Q-value networks are in synchronization.

The actor network may be modified in accordance with the policy gradient at each step for a single sample. The mini-batch has N_B transition samples, and the policy gradient may be computed as [48]:

$$\nabla J = \frac{1}{N_B} \sum \nabla_{\theta^{critic}} Q(s_t, a_t|\theta^{critic}) \nabla_{\theta^{actor}} \pi(st|\theta^{actor}) \quad (13)$$

where θ^{actor}, θ^{critic} represents the parameters of the function of loss, which can be described as the squared error among the target and the training.

The gradient of the target critic network is represented as $\nabla_{\theta^{critic}} Q(s_t, a_t|\theta^{critic})$, while $\nabla_{\theta^{\text{actor}}} \pi(st|\theta^{\text{actor}})$ represents the gradient of the training actor network under the parameter of θ^{actor}.

For critic network training, the Q-value with regard to long-term reward has been defined by the Bellman equation. The target Q-value is produced by inputting the target actor network's output in accordance with state s_{t+1} can be expressed as:

$$y = R_t\,(s_t, a_t) + \beta max(\theta^{target}|s_{t+1}, a_{t+1}) \quad (14)$$

where β is the discount factor.

Lastly, the loss function is minimized to update the assessment critic network:

$$L = \frac{1}{N_B} \sum_{t=1}^{N_B} (y - \theta^{train}|s_t, a_t)^2 \quad (15)$$

With the use of soft updating, the target network is going to periodically update the network weights from the main network for both actors and critics [47]. To construct the policy while avoiding the unstable and divergent trend appears in Q-learning.

The following gradient rule's updates on the target critic network and the target actor network are given as follows:

$$\theta^{target} = \tau_a \theta^{train} + (1 - \tau_a)\theta^{target} \quad (16)$$

$$\theta^{train} = \tau_c \theta^{target} + (1 - \tau_c)\theta^{train} \quad (17)$$

where τ_a and $\tau_c \ll 1$ are the learning rates for the target critic network's soft updating coefficient and the target actor-network's soft updating coefficient, respectively.

It's important to notice that this update strategy involves slowly updating the target network parameters while tracking the assessment network that has been learned. The total framework for the DDPG algorithm is shown in Figure (3), where the actor and critic-networks use various structures, respectively.

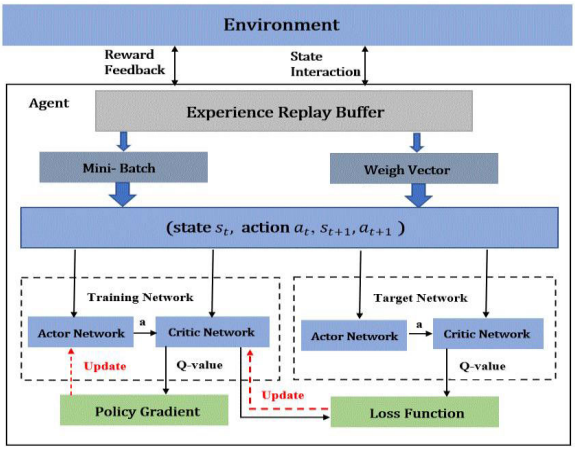

FIGURE 3. Flow diagram of the DDPG proposed algorithm.

2) DDPG ALGORITHM PROCESSING FRAMEWORK

The DRL proposed framework consists of a state set s_t, an action set a_t, a reward set r_t, and an agent IRS-UAV, which executes a certain action to obtain the proper reward while updating the current states. The actions are going to be employed to receive better rewards in various environments and reinforced iteratively. Moreover, since the THz band IRS phase shifts are thought to be continuous factors with very high quantization levels, it is motivated to use DDPG to build a two-layered actor-critic network in order to solve the issue with continuous solutions.

According to all of the aforementioned discussions, Algorithm 2 is a summary that outlines the DDPG's proposed algorithm in detail.

The IRS UAV is used as an agent by the DDPG algorithm in the THz network environment. The remaining corresponding elements, the state, action, and rewards, are clearly described as follows:

a: STATE SPACE

The set of states, contains the various observations that represent the environment. The state $\mathbf{s_t}$ of time step t is described as:

$$\boldsymbol{s_t} = [\theta_1^{(t-1)}, \ldots \theta_{2N}^{(t-1)}, x^{(t-1)}, y^{(t-1)}] \tag{18}$$

where $[\theta_1^{(t-1)}, \ldots .\theta_{2N}^{(t-1)}]$ denotes the IRS phase shift and $[x^{(t-1)}, y^{(t-1)}]$ represents the UAV optimal location at time $t-1$.

b: ACTION SPACE

The action space is made up of the choices that are available to an agent when transitioning from current state to the next state.

The action that the agent takes in time t, which includes IRS phase shift and UAV motion, entirely corresponds with the two optimization parameters:

- Deployments in the optimal position: For the deployment, the agent selects the UAV agent's next move for every t. The proposed approach enables the agent to identify the optimal movement at every moment while keeping the long-term reward in mind.

- IRS Phase Shifts: For every element expressed in the system, the agent determines the optimal phase shift for the time being.

Neglecting the time necessary to rotate the angle of the reflecting elements. Thus, the action function is defined as:

$$\boldsymbol{a_t} = \left[\theta_1^{(t)}, \ldots .\theta_{2N}^{(t)}, x^{(t)}, y^{(t)}\right] \tag{19}$$

The agent inputs the state s_t at time step t to obtain the corresponding action based on the current environment.

The agent then gets the optimal horizontal location q and the updated IRS phase shift.

Algorithm 2 Proposed DDPG Based Algorithm

Initialization: Randomly initialize the critic evaluation network $Q(s_t, a_t|\theta^{\text{train}})$ and the actor evaluation network $Q(s_t, a_t|\theta^{\text{target}})$ with their corresponding parameters θ^{target} and θ^{train} Initialize the experience replay buffer C, discount factor β, soft update coefficient τ and the minibatch size N_B
for episode $j = 1, \ldots, J$ ***do***
  Obtain the initial observed state s_t(18)
  Initialize the random process for action exploration
  for step $t = 1, \ldots, T$ **do** (for each step in episode)
    Select action a_{t+1} from the actor network
    Extract the actions at and observe new state s_{t+1} and rewards r_{t+1}
    Store transition s_t, a_t, r_t, s_{t+1} into the replay buffer C
    Sample the random N_B minibatch transitions from C to train
    Calculate the target Q-value by equation (14)
    Update the actor network by using sampled policy gradient in (13)
    Update the critic network by minimizing the loss function (15)
    Update two target networks via soft updates (16) and (17)
  end for
end for

c: REWARD

The agent obtains a reward $r_t(s_t, a_t)$ after performing action a_t in state s_t at time t.

In keeping with our goal, describe the reward as the sum data rate per group:

$$\boldsymbol{r_t} : R_{sum}^{(t)} = \sum_{m=1}^{M} R_k^{(t)}, m = 1, \ldots, M \tag{20}$$

V. RESULTS OF SIMULATION AND PERFORMANCE EVALUATION

This section, the performance of the proposed algorithm is evaluated for a scenario in an outdoor environment in using extensive simulations. The simulation is executed in the

TABLE 1. Simulation parameters for system environment settings.

Definition	Parameters	Values
Separations between reflectors	$\delta_x\ \delta_y$	0.1 m
Number of randomly distributed users UE	m	60
UAV altitude	h	55 m
Carrier Frequency	f	0.4 THz
Absorption coefficient	$k(f)$	0.005
Bandwidth	B	0.1 THz
Max. transmit power	$P_{\max}$	5 W
Minimum rate	R^{min}	0.02 Tbps
Noise power spectral density	$\sigma 2$	-174 dBm/Hz

TABLE 2. Simulation parameters for network model settings [49].

Definition	Parameters	Values
Episodes number	J	6000
Each episode's number of steps	T	10000
Discount factor	β	0.99
Hidden layers	Hl	2
Rate of learning for training the critic network	θ^c	0.001
Learning Rate of training actor network	θ^a	0.001
Learning Rate of target critic network	τ_c	0.001
Learning Rate of target actor network	τ_a	0.001
Size of replay buffer	C	200000
Size of sampled mini-batch	N_B	120

Python programming language. Both the actor-network and the critic-network in the proposed DDPG algorithm use the same fully connected neural network structure that includes an input layer for status information, an output layer that outputs the optimal action, two layers that are hidden in this simulation using the Rectified Linear Unit (ReLU) function as the function of activation, and modular normalization components, as shown in Figure (3). Additionally, Adam Optimizers should be used to train the policy network and the Q-network, and the parameters should be initialized at random using the zero-mean normal distribution.

It is worth mentioning that the simulation results presented were averaged over 500 independent iterations.

Moreover, we consider the communication scenario in an urban area of 100 m2, including one BS in the middle of the area serving 60 ground users. The positions of the users are uniformly distributed inside a circle area with a coverage radius of r = 6 m. Each user's total input data rate ranges from [250] to [550] Mbps. The users' positions are assumed to be fixed in each episode. Fly-IRS assumed hovering at a fixed altitude of $h_i = 55$ m and move with a velocity of $vc =$ 15 (m/s), other parameters are shown in Table 1 [33].

The simulation environment settings are illustrated in Table 1, and the simulation hyperparameters of the network model for the suggested DDPG algorithm with the actor and critic structures used in Algorithm 2 are displayed in Table 2 [49].

To illustrate the effectiveness of the proposed framework and compare its performance with the following three benchmark schemes that have been specified as follows:

A. THz-ENABLED UAVs SCHEME

This is the case where the BS provides communication to all users without needing the assistance of the IRS configuration. That is an approach to maximizing the sum data rate by only serving the BS to all users without IRS in the environment [33].

B. LOCAL SEARCH PHASE SHIFT SCHEME

This scheme is chosen an optimal value of IRS phase shift when the SINR constraint is satisfied. clearly, for each IRS element, all possible values are traversed. After that, this optimal value is utilized as the new value of IRS element to optimize another phase shift, until all phase shifts are fully optimized [34].

C. PRIMAL-DUAL PPO SCHEME

This scheme is employed to compare and assess the performance of the IRS model. As with the traditional IRS that is placed on a building, consider phase shift optimization. Herein, the IRS phase shift is optimized by the primal-dual proximal policy optimization algorithm [35].

D. RANDOM PHASE SHIFT

The actions in this scheme are picked by randomly selecting the IRS phase shift reflecting matrices, and the phase shifts of the IRS elements in this case have been determined with randomly configured values. This scheme's performance is referred to as the actual lower bound since the matrices are selected at random.

It is discovered that by employing the proposed DDPG algorithm, the rewards may be constantly enhanced and convergence to a constant value round episode may occur in both the non-IRS-assisted and IRS-assisted instances.

The next Figures display the average data rate (in Gbps) with the benefits of the IRS elements and evaluate the efficacy of the proposed flying Fly-IRS algorithm with optimal UAV location and optimized phase shift. Also, compare the proposed DDPG algorithm's performance with these schemes: Primal-dual PPO scheme, THz-enabled UAVs scheme, the Local Search phase shift scheme, the Random Phase Shift, and the proposed algorithm Without IRS.

Figure (4) illustrates the average sum data rate versus transmission power. Take a look at two cases of system

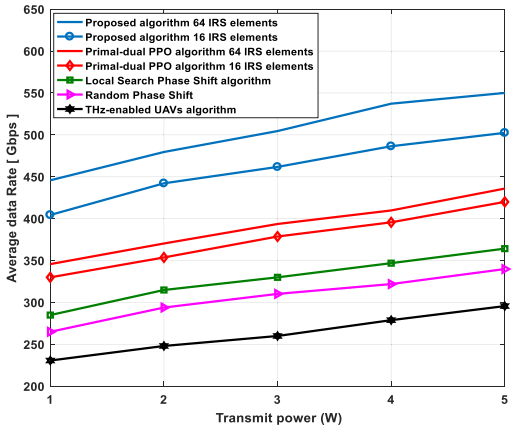

FIGURE 4. The average sum data rate versus different total transmit power.

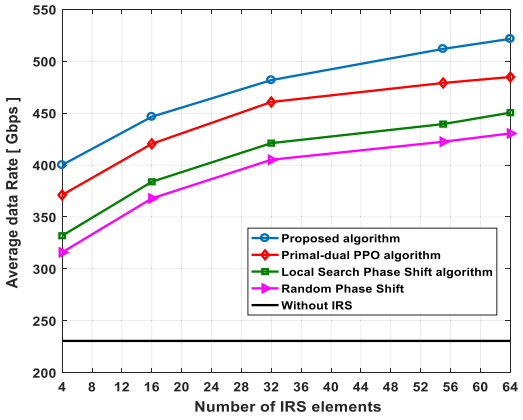

FIGURE 5. The average sum data rate versus the number of IRS elements.

parameter setups, one with IRS elements N = 16 and the other with N = 64.

In all schemes, it has been observed that increasing the power of transmission at the BS results in a higher overall average data rate. The proposed algorithm outperforms other compared techniques, with a larger performance gap as transmit power increases due to the more stringent constraints imposed by problem ***P1***. For example, when using 64 IRS elements, the sum data rate is around 540 Gbps, while 4 Watts of transmit power are needed when the proposed DDPG algorithm is deployed. Therefore, compared with the other approaches, we reach 409 Gbps in the primal-dual PPO scheme, 346 Gbps at the Local Search Phase Shift scheme, the Random Phase Shift case is about 321 Gbps, and the THz-enabled UAVs scheme can achieve 278 Gbps. Moreover, it may be observed that the system data rate steadily increases and tends to stabilize progressively. The explanation for this is that the average system data rate grows with the transmission maximum power limitation. Additionally, because channel interferences can't be disregarded under high $P_{\max}$, the system rate of data is going to eventually reach convergence.

Here, the figures depict the impact of IRS size, and the main observations are summarized below. In Figure (5), the average data rate is determined as a function of the number of IRS elements with the various comparison schemes mentioned.

The performance of the average data rate aided-THz communications is also significantly impacted by the number of IRS elements reflecting. It may be observed that the proposed algorithm, with the optimized UAV location and IRS phase shift, can achieve a data rate of 450 Gbps when there are 16 IRS elements and an approximate 530 Gbps sum data rate when employing IRS with 64 elements. By providing the IRS elements, the sum data rate gap between the proposed model and the Primal-dual PPO case increases. When the IRS phase shift is optimized, the performance difference between the proposed model and the primal-dual PPO models becomes even more noticeable. Furthermore, the data rate of the other techniques is about 485 Gbps at the primal-dual PPO algorithm, the Local Search Phase Shift scheme is about 450 Gbps, 430 Gbps at the Random Phase Shift case, and reach 230 Gbps at Without IRS for 64 elements. This is because the Fly-IRS may successfully change the phase of reflected signals to enhance the quality of the received signal, allowing the desired THz wireless channel to enhance a perfect line-of-sight. By improving UAV location and IRS phase shift, the proposed algorithm may attain an optimal balance between the BS and users, as well as improve the channels of all links, thereby improving the system performance.

In addition, it may be shown that the average data rate increases as the IRS increase when compared to other schemes, this is because of the large number of elements in the IRS employed, which leads to a higher gain in the system. This is not the case for the method Without IRS, which isn't impacted by the number of IRS elements. There is only a slight gain when compared to the Random phase shift approach. Additionally, increasing the number of IRS elements is an effective way to improve the sum data rate. However, this also results in a larger size of training data and more neurons, which increases complexity and output latency. Therefore, when constructing an IRS system, it is crucial to take into account the tradeoff between sum data rate and complexity.

Figure (6) displays the average sum data rate against the number of users, where N = 64 IRS elements. We additionally show an instance in which the IRS is configured at random, the case without the IRS, and the primal-dual PPO algorithm for the comparison. In all four scenarios, it is observed that as the user number rises, the sum data rate initially rises and then falls. The proposed DDPG approach surpasses the other cases and achieves the best performance for all scenarios.

Another observation from this Figure is that once the number of users is 50, the proposed algorithm may achieve a sum data rate of 540 Gbps. Hence, compared with the other approaches, we reach 525 Gbps in the primal-dual PPO algorithm, 512 Gbps at the Local Search Phase Shift scheme, 506 Gbps in the Random Phase Shift case, and the

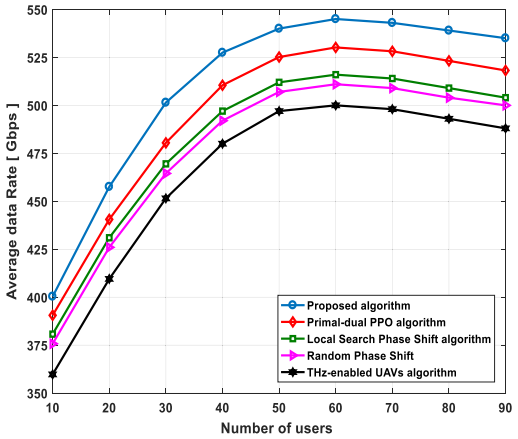

FIGURE 6. Average sum data rate versus the number of users.

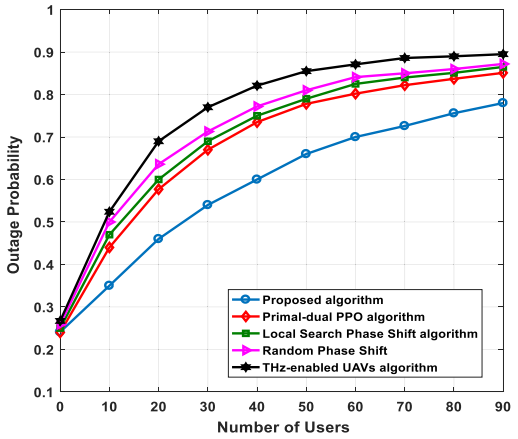

FIGURE 7. Outage probability versus the number of users.

THz-enabled UAVs scheme can achieve 496 Gbps. This is due to the fact that the signal-to-noise ratio is significantly greater at THz frequencies because of the high path-loss characteristics of terahertz channels, leading to little user interference. This demonstrates that the carefully configured IRS outperforms the proposed algorithm's gain over the scenario with a randomly configured IRS. Furthermore, once the number of users exceeds 70, user interference rises, and so the proposed DDPG works well in this scenario and delivers a greater sum rate of data compared to the other schemes. This is owing to the fact that the system capacity is considerably degraded at the THz frequency since the signal-to-noise ratio is significantly greater owing to the high pathloss property of the terahertz channel and low user interference.

The Outage Probability (OP) here can be described in the proposed case as the percentage of unsatisfied users whose downlink an occurrence of achievable data rate is below a required predefined data rate settled. (The application operating in THz network the range data rate of each user is between [250, 550] Mbps. So, the average threshold required data rate value) denoted by γ_{th} which is set to be 325 Mbps.

The P_i^{OP} of the THz communication under consideration may be calculated as [50]:

$$P_i^{OP} = P(R < \gamma_{th}) \tag{21}$$

Figure (7) depicts the effect of increasing the number of users on the outage probability with different schemes. The simulation parameters state that the power P max is set to 4W when employing 64 IRS elements. As the number of users increases, the reason for this the network architecture rapidly changes, causing communication to become unstable and the probability of outages to increase. The figure illustrates that our proposed algorithm maintains the outage probability and obtains the minimum value in comparison with the other techniques. As a result, in our proposed algorithm, we take into account the optimized IRS phase shifts as well as the UAV location that leads to the best performance. This best performance is then followed by the Primal-dual PPO algorithm, and then comes the ''Local Search Phase Shift'' algorithm, ''Random Phase Shift'' scheme where only the IRS elements phase shifts are optimized randomly, owing to increase in the outage probability, particularly with the increase in the number of users. The highest OP is achieved by the ''THz-enabled UAVs'' algorithm, where no IRS assists in the communication. This is because of increasing system demands, particularly when the number of users increase, as the rise increases the probability of users' links failing. That might cause the achievable rate of data to decline, which would therefore have an impact on the probability of an outage.

The data satisfaction rate ratio, due to the usage of the probability of satisfying users or not in outage probability, is used to determine the percentage of the links that satisfy the data rate threshold R_{th}.

Therefore, can be calculated for each link as [51]:

$$\mu_m = \frac{1}{1 + e^{-\delta(\frac{R_{total}^{e2e}}{R_{th}})}} \tag{22}$$

It can be found that the normal sigmoid function is used to determine the satisfaction rate, which is more suitable to show the rate between [0, 1]. A user's level of satisfaction may be employed to determine if or not they are satisfied.

Figure (8) demonstrates the user data satisfaction rate versus the different number of users under 64 IRS elements, $P_{\max} = 4$ watts. The user is stated to be satisfied if their satisfaction level is greater than or equal to the average threshold of 325 Mbps, and vice versa. It may be observed clearly from Figure (8) that the user data satisfaction rate decreases monotonically with an increased number of users because the users are going to be more satisfied if they're receiving a greater data rate above their minimum demand. This implies that the rate of data satisfaction among users improves when the number of IRS elements grows, due to the IRS ability to effectively change the phase of reflected signals in order to improve the quality of received signal, and then SINR increased in equation (9) could obtain better channel gain. Our analytical results are in good agreement with the

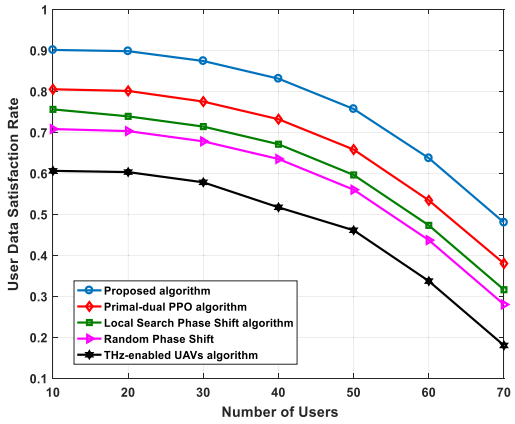

FIGURE 8. Satisfaction rate against increase of the number of users.

TABLE 3. Capacity of the system with numbers of IRS elements.

	Proposed Algorithm				Primal-dual PPO algorithm				THz-Enabled UAVs Algorithm	Local Search Phase Shift Algorithm			
	Number of IRS Elements				Number of IRS Elements				Number of IRS Elements	Number of IRS Elements			
	4	16	32	64	4	16	32	64	Without IRS	4	16	32	64
Capacity of the System	144	204	237	255	105	156	189	210	159	95	142	178	200

simulations, and the obvious scheme performs enhanced than the benchmark schemes.

Table 3 summarizes the results of comparisons of system capacity performance of four schemes with different numbers of IRS elements. This Table illustrates the effect of the increasing number of IRS elements on the system capacity at the average needed data rate of 325 Mbps, assuming that the number of users is M = 300 and employing IRS elements with N = 4, 16, 36, and 64. As it is demonstrated, the table also presents how the number of IRS reflecting elements increase with system capacity. This is because we designed an optimal IRS phase shifting matrix for the optimal location of the UAV in our system. As seen, when the necessary data rate is greater than the threshold of 325 Mbps, each network may satisfy the requirements of nearly all users, escalating the satisfied-user ratio in the system. Based on the results shown in Table 3, the system capacity of our proposed algorithm using the DDPG algorithm outperforms the system capacity with the Primal-dual PPO algorithm and THz-enabled UAVs scheme, where it was successful in improving system capacity by an average of 90% over the system employing the THz-enabled UAVs scheme. This is due to the fact that this algorithm works without the assistance of the IRS element configuration.

On the other hand, comparing the primal-dual PPO algorithm to the DDPG proposed algorithm, the latter tends to provide better performance in satisfied user ratio boosting by 30% and 25%, 21% considering the use of 16, 32, 64 IRS elements respectively in the primal-dual PPO algorithm. However, for the Local Search Phase Shift Algorithm, the increase in the number of elements has caused the interference to increase significantly, which leads to a slower growth from 16 to 32 elements, and results in a decrease when N exceeds 64. Finally, comparing the proposed algorithm with the Local Search Phase Shift scheme, 43%, 33%, and 28% performance gain is obtained for N = 16, 32 and 64 elements respectively, for the required data rate of 325 Mbps.

Furthermore, the increase in system capacity slows down as the number of IRS reflecting elements becomes too large. This is due to the greater number of IRS reflecting elements have an impact on the system capacity performance in the IRS assisted-THz communications, while the power consumption provided by the extra elements is relatively high. The performance of system capacity may achieve a significant convergence value at the expense of the training rate size, which increases hardware complexity and output latency. Therefore, when constructing an IRS-UAV system, it is crucial to take into account the tradeoff between sum data rate and complexity.

Figure (9) shows the rate of user data satisfaction for various numbers of IRS elements when the number of users is 70. Additionally, it has been established that the user data satisfaction rate increases with the number of IRS elements and achieves a data satisfaction rate ratio of 90% when N = 64 elements in the proposed algorithm. This is due to we proposed cost-effective algorithm based DDPG approach to select the IRS phase shift while considering the consumed beamforming training time. We considered a nearly passive IRS phase shift which adopt a combination of some active reflecting elements, and passive reflecting elements in sequence, to reflect more signal paths, signal power, and IRS elements to improve SINR at the cost of a slight performance loss.

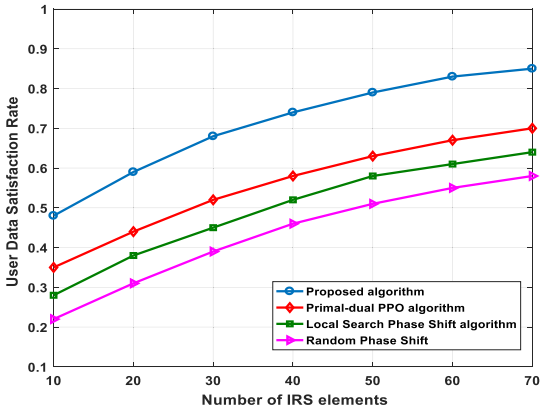

FIGURE 9. Performance of data satisfaction rate the versus number of IRS element.

Figure (10) clearly shows the episode reward of the DDPG proposed algorithm when employing IRS with N = 16, 24,

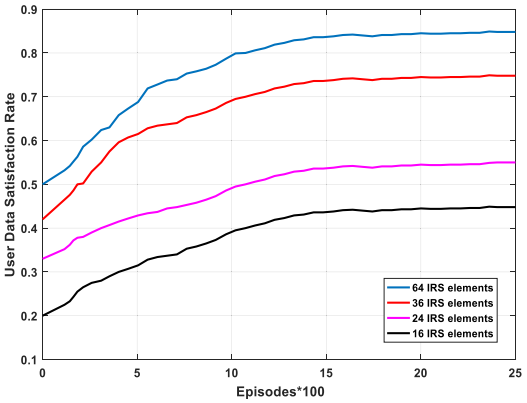

FIGURE 10. Performance of data satisfaction rate versus episodes, under different IRS elements for the proposed algorithm.

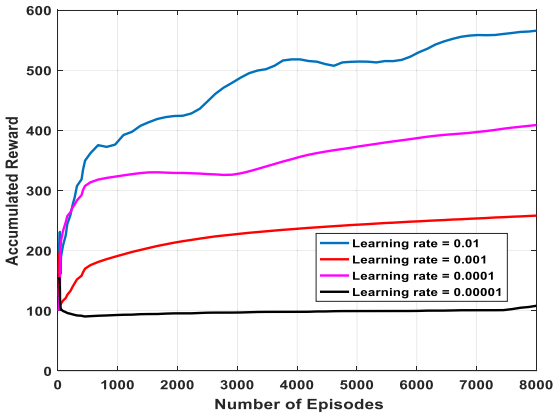

FIGURE 11. Accumulated rewards versus a function of episodes with different learning rates.

36, and 64 elements. It is apparent that as the number of IRS elements increases, the convergence value of the episode reward increases. The explanation for this is that IRS with more elements may be employed to offer more accurate phase shift rules to enhance the user data satisfaction rate, which efficiently enhances the instant reward. Otherwise, the user data rate could prove dissatisfaction, and hardware resources would be wasted.

Effect of the learning rate for our proposed critic and actor networks, we use constant learning rates. Figure (11) shows the average rewards for the critic and actor networks under various constant learning rates, i.e., [0.01, 0.001, 0.0001, 0.00001] as calculated by equations 15, 16, and illustrates the effects of these rates on convergence and rewards. It should be noted that the average rewards depend on the learning rate. It was observed that this updating technique entails slowly following the parameters network that has been taught in order to update the target network's parameters. In particular, the small setting (e.g., 0.00001) or the large setting (e.g., 0.001) both yield lesser average rewards, but the 0.01 rate of learning yields the best rewards. Better rewards, on the other hand, may result in increased convergence times. Since more rewards imply a bigger state space, it could take a longer time to converge on the best solution.

In conclusion, we may infer that DRL-DDPG is a complex learning process and that its performance may be influenced by a number of hyperparameters, particularly when the environment is rapidly changing. These hyperparameters also comprise the initialization and system settings, as well as the rate of learning, rate of decaying, minibatch size, and hyperparameters.

Additionally, it ought to be highlighted that proper DRL parameter adjustment and settings will significantly improve performance and reduce the convergence time.

VI. CHALLENGES AND FUTURE DIRECTIONS

Although Fly-IRS aided-THz communication proposed scheme based on DRL-DDPG has been investigated in several directions, it still encounters various difficulties for its applications. In this section, several main challenges for future our research are discussed.

A. FRAMEWORK CONSTRUCTION

The primary problem of the schemes is how to establish the appropriate DRL-DDPG architecture in the highly dynamic environment, with the help of Flying IRSs aided THz communication network is still a challenging and its complex. However, the challenges for the THz communications band are the THz channels are characterized by dynamic frame structure, which results in: large path loss hence limited network coverage, and visible-light-like propagation characteristics hence poor support of mobility in blockage-rich environments. In particular, for THz communication that is highly sensitive to blockage and pathloss. So, we must conduct a deeper study how to deploy the appropriate Flying IRS system with lower cost to improve the area coverage extension in the THz communication system.

B. TRAINING OVERHEAD AND COMPLEXITY

To coordinate multiple Flying IRSs to maximize the overall network performance, the DDPG architecture requires a large amount of communication channel data for building models, network training, and performance analysis. The performance of system capacity may achieve a significant convergence value at the expense of the training rate size, which increases hardware complexity and output latency. Moreover, to better meet the highly dynamic environment of Flying IRSs, the reward behavior and learning efficiency are contradictory objects in complex problems. So, we should analyze the input and output design and study of DDPG to guide the optimal allocation of control variables to cover larger data models.

VII. CONCLUSION

This paper proposes a framework for designing Flying IRS algorithm named a (Fly-IRS) aided-THz wireless network to optimize the system data rate and improve the system's performance. Accordingly, we formulated an optimization problem that by jointly optimizing the user Grouping and the

phase shift matrix of IRS with the optimal UAV's location to achieve the system data rate maximization, enhancing the system capacity, and minimizing the Outage Probability to provide a better satisfied user ratio. To solve the formulated problem, we have transformed it into two sub-problems. Consequently, an iterative algorithm based on modified K-means clustering algorithm is proposed to solve the first sub-problem the optimizing user Grouping, while Deep Deterministic Policy Gradient (DDPG) algorithms is proposed to be employed due to their low complexity, to optimize the IRS phase shift and optimum UAV's location optimization. The Simulation results demonstrate how to assess the proposed algorithm's performance to maximize the system's data rate by up to 95% and improve the capacity of the system on average by 94% tends to provide better performance in satisfying users, boosting the ratio by 90% compared with other benchmarks. Moreover, the system capacity performance is investigated to define the optimal number of IRS elements necessary for the system, and it is discovered that the optimal number of IRS elements needed is 64 elements in the proposed algorithm, escalating the user satisfaction ratio for this system. Future works, the flying IRS energy is also a matter of concern, with limited power of UAV. We would like to extend our work in the next step by optimized the UAV trajectory while minimizing the time consume of the system, reduce the total flying time of UAV, and minimizing the UAV propulsion energy.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.

[2] S. S. Hassan, Y. K. Tun, W. Saad, Z. Han, and C. S. Hong, "Blue data computation maximization in 6G space-air-sea non-terrestrial networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.

[3] S. Salman Hassan, D. Hyeon Kim, Y. Kyaw Tun, N. H. Tran, W. Saad, and C. Seon Hong, "Seamless and energy-efficient maritime coverage in coordinated 6G space–air–sea non-terrestrial networks," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 4749–4769, 2022.

[4] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8315–8328, Dec. 2019.

[5] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng, and L. Hanzo, "Hybrid precoding for wideband millimeter wave MIMO systems in the face of beam squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1847–1860, Mar. 2021.

[6] M. Polese, J. M. Jornet, T. Melodia, and M. Zorzi, "Toward end-to-end, full-stack 6G terahertz networks," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 48–54, Nov. 2020.

[7] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, and L. Hanzo, "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, 2nd Quart., 2023.

[8] L.-H. Shen, K.-T. Feng, and L. Hanzo, "Five facets of 6G: Research challenges and opportunities," *ACM Comput. Surv.*, vol. 55, no. 11, pp. 1–39, Nov. 2023.

[9] K. M. S. Huq, J. Rodriguez, and I. E. Otung, "3D network modeling for THz-enabled ultra-fast dense networks: A 6G perspective," *IEEE Commun. Standards Mag.*, vol. 5, no. 2, pp. 84–90, Jun. 2021.

[10] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, "An overview of signal processing techniques for terahertz communications," *Proc. IEEE*, vol. 109, no. 10, pp. 1628–1665, Oct. 2021.

[11] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, "Can terahertz provide high-rate reliable low-latency communications for wireless VR?" *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9712–9729, Jun. 2022.

[12] R. Pant and L. Malviya, "THz antennas design, developments, challenges, and applications: A review," *Int. J. Commun. Syst.*, vol. 36, no. 8, p. e5474, May 2023.

[13] L. Xu, M. Chen, M. Chen, Z. Yang, C. Chaccour, W. Saad, and C. S. Hong, "Joint location, bandwidth and power optimization for THz-enabled UAV communications," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 1984–1988, Jun. 2021.

[14] O. A. Amodu, C. Jarray, S. A. Busari, and M. Othman, "THz-enabled UAV communications: Motivations, results, applications, challenges, and future considerations," *Ad Hoc Netw.*, vol. 140, Mar. 2023, Art. no. 103073.

[15] F. Kiani, A. Seyyedabbasi, R. Aliyev, M. A. Shah, and M. U. Gulle, "3D path planning method for multi-UAVs inspired by grey wolf algorithms," *J. Internet Technol.*, vol. 22, no. 4, pp. 743–755, Jul. 2021.

[16] F. Kiani, A. Seyyedabbasi, R. Aliyev, M. U. Gulle, H. Basyildiz, and M. A. Shah, "Adapted-RRT: Novel hybrid method to solve three-dimensional path planning problem using sampling and metaheuristic-based algorithms," *Neural Comput. Appl.*, vol. 33, no. 22, pp. 15569–15599, Nov. 2021.

[17] X. Wang, P. Wang, M. Ding, Z. Lin, F. Lin, B. Vucetic, and L. Hanzo, "Performance analysis of terahertz unmanned aerial vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16330–16335, Dec. 2020.

[18] Y. Li, C. Yin, T. Do-Duy, A. Masaracchia, and T. Q. Duong, "Aerial reconfigurable intelligent surface-enabled URLLC UAV systems," *IEEE Access*, vol. 9, pp. 140248–140257, 2021.

[19] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795–202818, 2020.

[20] N. S. Perovic, L.-N. Tran, M. Di Renzo, and M. F. Flanagan, "Achievable rate optimization for MIMO systems with reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3865–3882, Jun. 2021.

[21] Y. Chen, D. Chen, and T. Jiang, "Beam-squint mitigating in reconfigurable intelligent surface aided wideband mmWave communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.

[22] M. Al-Jarrah, A. Al-Dweik, E. Alsusa, Y. Iraqi, and M.-S. Alouini, "On the performance of IRS-assisted multi-layer UAV communications with imperfect phase compensation," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8551–8568, Dec. 2021.

[23] H. Du, J. Zhang, K. Guan, D. Niyato, H. Jiao, Z. Wang, and T. Kürner, "Performance and optimization of reconfigurable intelligent surface aided THz communications," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3575–3593, May 2022.

[24] R. Aziz and T. Girici, "Deployment of a UAV-mounted intelligent reflecting surface in the THz band," in *Proc. Int. Balkan Conf. Commun. Netw. (BalkanCom)*, Aug. 2022, pp. 168–172.

[25] N. Abuzainab, M. Alrabeiah, A. Alkhateeb, and Y. E. Sagduyu, "Deep learning for THz drones with flying intelligent surfaces: Beam and handoff prediction," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6.

[26] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski, and M. Debbah, "Seven defining features of terahertz (THz) wireless systems: A fellowship of communication and sensing," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 967–993, 2nd Quart., 2022.

[27] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive MIMO with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1376–1393, Feb. 2021.

[28] M. Shehab, A. Almohamad, M. Elsayed, A. Badawy, T. Khattab, N. Zorba, M. Hasna, and D. Trinchero, "Terahertz multiple access: A deep reinforcement learning controlled multihop IRS topology," 2023, *arXiv:2303.09476*.

[29] K. Tao, Q. Wang, and D. Yue, "Data compression and damage evaluation of underground pipeline with musicalized sonar GMM," *IEEE Trans. Ind. Electron.*, vol. 99, pp. 1–9, 2023.

[30] Z. Yao, X. Liang, G.-P. Jiang, and J. Yao, "Model-based reinforcement learning control of electrohydraulic position servo systems," *IEEE/ASME Trans. Mechatronics*, vol. 28, no. 3, pp. 1446–1455, Jun. 2023.

[31] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Sum-rate maximization for intelligent reflecting surface assisted terahertz communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 3320–3325, Mar. 2022.

[32] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, and Z. Zhang, "Hybrid beamforming for RIS-empowered multi-hop terahertz communications: A DRL-based method," in *Proc. IEEE Globecom Workshops (GC Wkshps*, Dec. 2020, pp. 1–6.

[33] S. S. Hassan, Y. Min Park, Y. K. Tun, W. Saad, Z. Han, and C. S. Hong, "3TO: THz-enabled throughput and trajectory optimization of UAVs in 6G networks by proximal policy optimization deep reinforcement learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 5712–5718.

[34] S. Farrag, E. Maher, A. El-Mahdy, and F. Dressler, "Reconfigurable intelligent surface and UAV-assisted THz mobile communications," in *Proc. 27th Eur. Wireless Conf.*, Sep. 2022, pp. 1–8.

[35] X. Liu, H. Zhang, K. Long, M. Zhou, Y. Li, and H. V. Poor, "Proximal policy optimization-based transmit beamforming and phase-shift design in an IRS-aided ISAC system for the THz band," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2056–2069, Jul. 2022.

[36] K. Dovelos, S. D. Assimonis, H. Q. Ngo, B. Bellalta, and M. Matthaiou, "Intelligent reflecting surfaces at terahertz bands: Channel modeling and analysis," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6.

[37] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, Oct. 2011.

[38] Y. Zheng, S. Bi, Y. A. Zhang, X. Lin, and H. Wang, "Joint beamforming and power control for throughput maximization in IRS-assisted MISO WPCNs," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 8399–8410, May 2021.

[39] A. Ranjha and G. Kaddoum, "URLLC facilitated by mobile UAV relay and RIS: A joint design of passive beamforming, blocklength, and UAV positioning," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4618–4627, Mar. 2021.

[40] A. Trotta, M. Di Felice, F. Montori, K. R. Chowdhury, and L. Bononi, "Joint coverage, connectivity, and charging strategies for distributed UAV networks," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 883–900, Aug. 2018.

[41] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, Z. Zhang, and M. Debbah, "Multi-hop RIS-empowered terahertz communications: A DRL-based hybrid beamforming design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1663–1677, Jun. 2021.

[42] N. Casas, "Deep deterministic policy gradient for urban traffic light control," 2017, *arXiv:1703.09035*.

[43] F. H. O'Shea, N. Bruchon, and G. Gaio, "Policy gradient methods for free-electron laser and terahertz source optimization and stabilization at the FERMI free-electron laser at Elettra," *Phys. Rev. Accel. Beams*, vol. 23, no. 12, Dec. 2020.

[44] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "UAV-assisted and intelligent reflecting surfaces-supported terahertz communications," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1256–1260, Jun. 2021.

[45] A. Raza, U. Ijaz, M. K. Ishfaq, S. Ahmad, M. Liaqat, F. Anwar, A. Iqbal, and M. S. Sharif, "Intelligent reflecting surface-assisted terahertz communication towards B5G and 6G: State-of-the-art," *Microw. Opt. Technol. Lett.*, vol. 64, no. 5, pp. 858–866, May 2022.

[46] A. S. Abdalla and V. Marojevic, "DDPG learning for aerial RIS-assisted MU-MISO communications," in *Proc. IEEE 33rd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2022, pp. 701–706.

[47] Q. Liu, J. Wu, L. Hu, S. Bi, W. Ji, and R. Yang, "Optimal energy efficiency used DDPG in IRS-NOMA wireless communications," *Symmetry*, vol. 14, no. 5, p. 1018, May 2022.

[48] J. Lin, Y. Zout, X. Dong, S. Gong, D. T. Hoang, and D. Niyato, "Deep reinforcement learning for robust beamforming in IRS-assisted wireless communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[49] S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar, and N. Aydin, "Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases," *Comput. Biol. Chem.*, vol. 97, Apr. 2022, Art. no. 107619.

[50] Z. Song, J. Feng, Z. Shi, Q. Dou, G. Yang, Y. Li, and S. Ma, "Outage probability analysis of HARQ-aided terahertz communications," in *Proc. 13th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2021, pp. 1–6.

[51] G. Sun, K. Xiong, G. O. Boateng, G. Liu, and W. Jiang, "Resource slicing and customization in RAN with dueling deep Q-network," *J. Netw. Comput. Appl.*, vol. 157, May 2020, Art. no. 102573.

SHEREEN S. OMAR received the B.S. and M.S. degrees in electronics and communication engineering from Helwan University, Cairo, Egypt, in 2010 and 2016, respectively, where she is currently pursuing the Ph.D. degree in electronics and communication engineering with the Faculty of Engineering. She has experience in wireless network research. Her research interests include terahertz-band communications and 6G communication networks, unmanned aerial vehicle (UAV), intelligent reflecting surfaces (IRS), AI application in wireless communication, and machine learning.

AHMED M. ABD EL-HALEEM (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electronics and communications engineering from Helwan University, Cairo, Egypt, in 2001, 2006, and 2012, respectively. He is currently an Associate Professor with the Electronics and Communications Engineering Department, Faculty of Engineering, Helwan University. He is a member of a research team that awards several applied research projects funded by national and international funding agencies in the field of wireless communication, the Internet of Things (IoT) and its applications, and smart education systems. His current research interests include mobile/vehicular ad-hoc communication networks, 5G and 6G radio access networks, cognitive radio networking, device-to-device communication, the IoT, reconfigurable intelligent surface (RIS), and AI application in wireless communication. This includes mobility management techniques and routing schemes for mobile ad-hoc networks (MANET), security and secure routing algorithms for ad-hoc wireless networks, and routing protocols for cognitive radio communication networks. This also includes AI-based applications for education and wireless networks, the IoT communications, unmanned aerial vehicle (UAV) communications, RIS, and smart grid communication systems.

IBRAHIM I. IBRAHIM is a former member of the National Communications and Electronics Engineering Promotion Committee, the former Head of the Electronics, Communications and Computers Engineering Department, and a Professor of wireless communications with the Faculty of Engineering, Helwan University, Cairo, Egypt. His current research interests include massive MIMO, antenna design, 5G and 6G radio access networks wireless communications, heterogeneous cellular networks, satellite communications, the Internet of Things, visible light communications, cloud radio access networks, device-to-device communications and unmanned aerial vehicle (UAV) communications, and reconfigurable intelligent surface (RIS).

AMANY M. SALEH received the B.Sc., M.Sc., and Ph.D. degrees in electronics, communications, and computer engineering from the Faculty of Engineering, Helwan University, in 2011, 2016, and 2021, respectively. She is currently a Teacher with the Faculty of Engineering, Helwan University. Her research interests include massive MIMO, antenna design, mutual coupling, channel model, metasurfaces, and 6G mobile networks.