Hybrid Optimization for NOMA-Based Transmissive-RIS Mounted UAV Networks

Zain Ali, Muhammad Asif, Wali Ullah Khan, Abdelrahman Elfikky, *Graduate Student Member, IEEE*, Asim Ihsan, Manzoor Ahmed, Ali Ranjha, and Gautam Srivastava, *Senior Member, IEEE*

***Abstract*—In this work, we introduce a novel hybrid joint optimization framework specifically designed for enhancing the performance of consumer electronics in vehicular networks using a transmissive reconfigurable intelligent surface (T-RIS)-mounted uncrewed aerial vehicle (UAV) system. The UAV employs the non-orthogonal multiple access (NOMA) protocol to broadcast data to multiple ground devices, ensuring efficient communication. Our primary objective is to maximize the overall system sum rate while adhering to key constraints such as the rate requirements of ground devices, UAV battery capacity, and UAV coordinate boundaries. The optimization challenge of maximizing the system's sum rate is inherently non-convex and complex. To address this, we decompose the problem into manageable subproblems. The beamforming optimization problem is tackled using successive convex approximation and semi-definite programming techniques, allowing for effective handling of non-convexity. For power allocation, we employ the Lagrangian dual method along with the sub-gradient technique, ensuring optimal power distribution among devices. To optimize the UAV's location, we propose a dueling-based double deep reinforcement learning (D3RL) framework. This approach effectively combines all computed solutions, resulting in a comprehensive joint optimization strategy. Simulation results highlight the exceptional performance of the proposed framework. Specifically, optimizing the UAV's location leads to a substantial performance gain of up to 65.9% compared to a system where only beamforming and power allocation are optimized with the UAV positioned at the center of the service area. These findings underscore the potential of our framework in advancing consumer electronics connectivity in vehicular networks.**

***Index Terms*—Machine learning, non-orthogonal multiple access (NOMA), resource allocation, transmissive reconfigurable intelligent surface (T-RIS), uncrewed aerial vehicle (UAV).**

Received 20 July 2024; revised 4 December 2024; accepted 7 January 2025. Date of publication 13 January 2025; date of current version 14 August 2025. *(Corresponding author: Gautam Srivastava.)*

Zain Ali and Abdelrahman Elfikky are with the Electrical and Computer Engineering Department, Baskin School of Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: zainalihanan1@gmail.com; abdo.fikky2020@gmail.com).

Muhammad Asif is with the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 518000, China (e-mail: masif@ujs.edu.cn).

Wali Ullah Khan is with the Interdisciplinary Center for Security, Reliability and Trust, University of Luxembourg, 1855 Luxembourg City, Luxembourg (e-mail: waliullah.khan@uni.lu).

Asim Ihsan is with the Department of Engineering, University of Cambridge, CB3 0FA Cambridge, U.K. (e-mail: ai422@cam.ac.uk).

Manzoor Ahmed is with the School of Computer and Information Science and the Institute for AI Industrial Technology Research, Hubei Engineering University, Xiaogan 432000, China (e-mail: manzoor.achakzai@gmail.com).

Ali Ranjha is with the Electrical Engineering Department, École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada (e-mail: ali-nawaz.ranjha.1@ens.etsmtl.ca).

Gautam Srivastava is with the Department of Mathematics and Computer Science, Brandon University, Brandon, MB R7A 6A9, Canada, also with the Research Centre for Interneural Computing, China Medical University, Taichung 40402, Taiwan, and also with the Centre for Research Impact and Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura 140401, India (e-mail: srivastavag@brandonu.ca).

Digital Object Identifier 10.1109/TCE.2025.3528929

I. Introduction

RECENT advancements in communication systems have enabled vehicular networks to achieve higher efficiency, increased capacity, enhanced adaptability, and robust connectivity for vehicles. However, the future demands of communication networks require extended coverage, especially in remote or disaster-stricken regions [1]. To connect users globally, the use of spatial and aerial nodes has been proposed [2], [3]. This involves providing services to distant devices using spatial nodes like satellites or aerial nodes like unmanned aerial vehicles (UAVs) [4], [5]. UAVs serve as valuable assets, functioning either as base stations to connect ground devices with a backhaul node [6] or as aerial relays between satellites and ground devices [7]. Satellite-based connectivity is ideal for areas far from ground infrastructure, offering vast coverage to connect devices to distant base stations [8]. However, satellite communication faces significant path losses, resulting in less efficient systems. In less dense areas with devices in remote locations, such as rural settings or where sensors collect data, UAVs are preferred due to better channel conditions and lower power requirements, providing robust connectivity and improved energy efficiency. Additionally, in urban areas, large gatherings in specific locations can overburden the local communication infrastructure. In such cases, a temporary UAV-based network can be deployed to facilitate users and reduce the load on the terrestrial network. Moreover, the number and positioning of UAV nodes can be adjusted to meet the communication needs of consumer devices, offering a dynamic and responsive solution compared to the fixed nature of satellites, which are either in continuous motion (low or medium-earth orbit) or stationary (geostationary) with respect to the Earth [9]. This flexibility in UAV deployment caters to diverse communication requirements across various geographical and operational contexts.

To harness the advantages of location optimization in UAV networks, an algorithm for optimizing the position of a single-antenna UAV was proposed by [10]. The single antenna assumption at each node ensures that the angle-of-departure at the UAV has no impact on the channel. Another study by [11]

introduced a location optimization scheme for a UAV network transmitting data to a single ground device, showcasing significant performance gains through UAV location optimization. Scaling up to a multi-user and multi-UAV scenario, [12] proposed a power allocation and location optimization scheme for UAV-assisted communication. Addressing the UAV location and user association problem in systems with multiple UAVs serving ground devices in an allocated region, [13] presented a comprehensive approach considering both uplink and downlink transmissions. Notably, all these works consider frequency division multiple access-based (FDMA) transmission, allocating a dedicated channel to each user to prevent interference.

Looking forward, future communication systems demand innovative solutions to meet the escalating requirements for higher spectral efficiency. Non-orthogonal multiple access (NOMA) introduces a groundbreaking approach, enabling simultaneous data transmission to multiple devices on the same frequency and time resources [14]. This advancement significantly enhances spectral efficiency, fostering more robust connectivity in dense networks [15]. Exploring NOMA transmission with UAVs, [16] proposed a power allocation optimization framework at the UAV, enhancing system efficiency. However, this study did not address UAV location optimization. Subsequently, [17] optimized both location and power allocation for a NOMA-UAV serving multiple ground stations, utilizing a convex solver for power allocation. In a similar vein, the work in [18] optimized both location and provided a closed-form expression for power allocation in NOMA-UAV scenarios. However, a notable drawback surfaced: the complexity of the proposed closed-form expression increases with the number of users in the system, inherent to NOMA systems. Moreover, in NOMA systems, allocating the same channel to a large number of users results in significant degradation of channel capacity, primarily due to the elevated interference levels within the system.

To amplify channel capacity, the reconfigurable intelligent surface (RIS) emerges as a promising candidate for future communication networks [19], [20]. By manipulating both signal amplitude and phase, RIS improves overall system performance [21]. With the capability to reflect or transmit the incident signals [22], [23], RIS effectively mitigates issues related to noise amplification and additional noise introduced by conventional relaying systems. Numerous studies explored the applications of reflective-RIS (R-RIS) in UAV networks. For instance, Khan et al. have considered R-RIS in NOMA device-to-device communication underlaying cellular networks [24]. In [20], the authors have used R-RIS to improve the physical layer security of wireless systems under eavesdropping and jamming attacks. The authors in [25] optimized the trajectory of an R-RIS mounted UAV, enhancing system efficiency. In a similar context, [26] proposed a Deep Reinforcement Learning (DRL) framework for optimizing passive beamforming at the R-RIS and UAV trajectory. Nevertheless, the proposed online training framework demonstrated slow convergence, requiring thousands of iterations, unsuitable for highly dynamic systems. The work in [27] proposed a DRL scheme for beamforming and location optimization of R-RIS-mounted UAVs, yet due to its online training model, this scheme demands numerous iterations and is sensitive to changes in network dynamics.

Moreover, the RIS can also function as a transmitter without necessitating intricate signal processing, setting it apart from conventional multi-antenna systems that rely on complex RF modules, thereby incurring high hardware costs [28], [29]. The transmissive-RIS (T-RIS) presents distinct advantages over the R-RIS for the following reasons: (a) R-RIS experiences self-interference as the feed antenna and receiver coexist on the same side, leading to interference between the incident and reflected signals, whereas in the case of T-RIS the feed antenna and the receivers are on the opposite side of the T-RIS; (b) In contrast to R-RIS, T-RIS can be designed with enhanced operational bandwidth and aperture efficiency [30]. Consequently, the integration of T-RIS-based UAV communication holds the potential to significantly enhance capacity, energy efficiency, quality of service, and spectrum utilization in emerging wireless communication systems. In [31], authors showed that incorporating T-RIS significantly enhances the coverage area of communication networks. Subsequent works, [32] and [33], optimize beamforming and power allocation to maximize the sum rate in downlink multi-user T-RIS systems, considering both downlink and uplink scenarios. However, these works do not explore the advantages of incorporating T-RIS into UAV systems, leaving the potential benefits of T-RIS-mounted UAVs in communication networks unexplored.

When a UAV is equipped with a single antenna and operates in FDMA style transmission, optimizing its location is straightforward, as observed in [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]. However, in the case of an RIS-mounted UAV, the angle of departure introduces complexity to location optimization. Additionally, integrating NOMA into a T-RIS-mounted UAV can enhance communication network performance but increases system complexity due to the interference introduced by multiple users sharing the same channel. Moreover, all the previous works assumed a uniform distribution of users, which is an idealistic and impractical assumption. To address a more realistic scenario, [34] optimized cell partitioning for UAV-aided communication networks with hotspot areas for ground devices. Similarly, [35] optimized both cell partitioning and bandwidth allocation for UAV communication systems considering the presence of ground user hotspots.

Moreover, it is worth noting that existing machine learning-based solutions for location optimization in RIS-mounted UAVs are susceptible to changing network conditions and require frequent retraining. For example, the solution in [26] doesn't account for the current locations of ground devices when determining the optimal UAV location, necessitating retraining whenever ground device locations change. Similarly, the DRL model in [27] uses the current coordinates of ground devices as distinct inputs to the DRL agent, demanding retraining when the number of users changes due to the change in the size of DRL state parameters.

A communication network featuring a T-RIS-mounted UAV employing NOMA for transmission holds the promise of

superior communication performance. However, the intricate challenge of optimizing resource allocation in this network remains unexplored in the existing literature. This paper introduces a hybrid framework designed for the joint optimization of beamforming, power allocation, and location for a NOMA-enabled T-RIS-mounted UAV. The proposed innovative approach eliminates the need for retraining the advanced DRL model when the number or location of ground devices changes, providing an efficient solution for T-RIS beamforming and NOMA power allocation. The problem of optimizing resource allocation in this system has not been explored in the literature before. The main contributions of the work are summarized as follows:

- We consider a system where a NOMA-enabled, T-RIS-mounted UAV serves ground devices by updating its location based on their positions and requirements.
- We optimize the UAV's location, T-RIS beamforming, and NOMA power allocation to maximize the system's sum rate, subject to constraints on the UAV's power budget, service area boundaries, ground device rate requirements, and beamforming.
- We used successive convex approximation and semi-definite programming for T-RIS beamforming and a Lagrangian dual with a sub-gradient method for optimal NOMA power allocation. For UAV location optimization, we introduced a dueling-based double deep reinforcement learning (D3RL) framework, which adapts to changes in device location or number without retraining. This hybrid approach combines the D3RL framework with beamforming and power allocation schemes for a joint solution.
- Simulations focused on hotspot areas of ground devices within the UAV-allocated region, reflecting practical device distributions [34], [35]. Results show the proposed scheme performs excellently, with the hybrid framework also achieving strong results for uniformly distributed ground devices.

The paper is structured as follows: Section II introduces the considered system model. The problem formulation is presented in Section III. Section IV explains the proposed solution framework. Simulation results and discussions are provided in Section V. Finally, Section VI concludes the work.

II. System Model

We consider a system where T-RIS-mounted UAVs are deployed to extend coverage to ground devices in remote or disaster-stricken regions. Each UAV is equipped with the same technology, tasked with servicing devices within its designated area. Employing the NOMA framework, UAVs broadcast data to ground devices. To mitigate interference from adjacent cells, we assume that UAVs in neighboring cells use different transmission frequencies. Consequently, we focus on optimizing resource allocation and location for a single UAV serving a cell, with the understanding that the same solution can be applied to all other cells in the system, thereby optimizing overall system performance.

In our considered system, each UAV is equipped with a T-RIS containing K_x elements in each column and K_y elements in each row, resulting in a total of $K = K_x \times K_y$ elements. Serving N users in the system, as illustrated in Fig. 1, and acknowledging the strong line-of-sight component in air-to-ground communication, we model the channels between the UAV and ground devices as Rician distributed. The channel gain from the UAV to the nth user is given by:

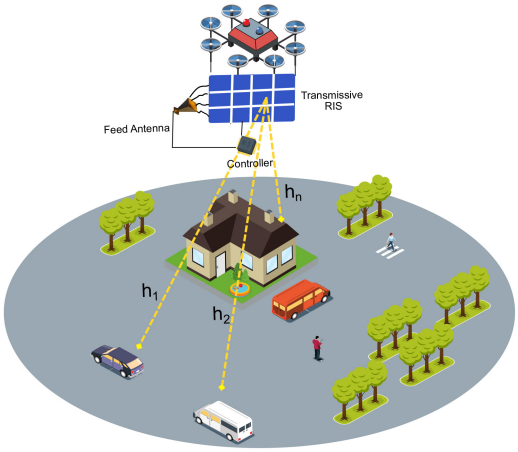

Fig. 1. T-RIS-assisted UAV-NOMA network.

$$h_n = \sqrt{\gamma_n}\left(\sqrt{\frac{K_n}{K_n+1}}h_n^{LoS} + \sqrt{\frac{1}{K_n+1}}h_n^{NLoS}\right), \quad (1)$$

where K_n is the Rician factor from the UAV to the nth user, $\gamma_n = \frac{\gamma_0}{d_n^{\varrho}}$, with γ_0 being the channel power at a reference distance of 1 meter. Here, d_n represents the distance of the nth user from the UAV, and ϱ is the path loss exponent. The distance d_n is computed as $d_n = \sqrt{(q_n^z - u^z)^2 + (q_n^x - u^x)^2 + (q_n^y - u^y)^2}$, where the location of the UAV is given by $\{u^x, u^y, u^z\}$, in the x, y and z coordinates, and the location of the nth device is given by $\{q_n^x, q_n^y, q_n^z\}$. Subsequently, h_n^{LoS} represents the line-of-sight component of the channel and is given by:

$$h_n^{\text{LoS}} = \left[1, e^{-j\beta \sin\theta_n \cos\phi_n}, \ldots, e^{-j\beta \sin\theta_n \cos\phi_n (K_x-1)}\right]^T \otimes \left[1, e^{-j\beta \sin\theta_n \sin\phi_n}, \ldots, e^{-j\beta \sin\theta_n \sin\phi_n (K_y-1)}\right]^T, \quad (2)$$

for $\beta = \frac{2\pi f_c \Omega}{c}$, where Ω is the distance between two adjacent T-RIS elements, f_c is the carrier frequency, c is the speed of light and $\otimes$ denotes the Kronecker product. The values of $\sin\theta_n$, $\sin\phi_n$, and $\cos\phi_n$ (where θ_n and ϕ_n denote the vertical and horizontal angle of departure to the nth ground device) are computed as [25]:

$$\sin\theta_n = \left(\frac{q_n^z - u^z}{d_n}\right), \quad (3)$$

$$\sin\phi_n = \left(\frac{q_n^x - u^x}{\sqrt{(q_n^x - u^x)^2 + (q_n^y - u^y)^2}}\right), \quad (4)$$

$$\cos\phi_n = \left(\frac{q_n^y - u^y}{\sqrt{(q_n^x - u^x)^2 + (q_n^y - u^y)^2}}\right). \quad (5)$$

Then, the non-line-of-sight component is considered to be independently and identically distributed as: $h_n^{NLoS} \sim \mathcal{CN}(0,1)$.

The beamforming vector at the T-RIS is denoted as $f = [\alpha_1 e^{j\psi_1}, \alpha_2 e^{j\psi_2}, \ldots, \alpha_K e^{j\psi_K}]^T$, where $\alpha_k \in [0,1]$ and $e^{j\psi_k}$ denotes the amplitude and phase of the signal transmitted by the kth RIS element. Without loss of generality, assuming the ground devices are sorted in descending order of channel gains at the UAV, the signal received by the nth device after applying successive interference cancellation is given by[1]:

$$y_n = h_n^{\dagger} f \sqrt{p_n} s_n + h_n^{\dagger} f \sum_{i=1}^{n-1} \sqrt{p_i} s_i + \chi_n, \tag{6}$$

where $\dagger$ is the Hermitian operator, p_n is the power allocated to the signal of the nth device, s_n is the symbol of the nth device and χ_n denotes additive white Gaussian noise with 0 mean and variance σ^2.

III. Problem Formulation

Considering that the UAV flies at a predetermined height, we aim to optimize the location, RIS beamforming, and the power allocation at the UAV to maximize the sum rate of the system. The objective of maximizing the sum rate is given as:

$$\max_{f,p_n,u_x,u_y} \sum_{n=1}^{N} B \log_2 \left(1 + \frac{p_n |h_n^{\dagger} f|^2}{|h_n^{\dagger} f|^2 \sum_{i=1}^{n-1} p_i + \sigma^2}\right), \tag{7}$$

where B is the bandwidth of the channel, f is the beamforming vector, and p_n is the transmission power allocated to the signal of the nth device, as defined before. The location parameters u_x and u_y affect the channel gains from the UAV to the ground devices, as explained in the previous section. The power budget constraint at the UAV is given as:

$$\sum_{n=1}^{N} p_n \leq P_{max}, \tag{8}$$

where P_{max} is the maximum allowed transmission power at the UAV. To satisfy the rate requirement of each device in the system, we introduce the following constraint:

$$B \log_2 \left(1 + \frac{p_n |h_n^{\dagger} f|^2}{|h_i^{\dagger} f|^2 \sum_{i=1}^{n-1} p_i + \sigma^2}\right) \geq R_{min}, \forall n, \tag{9}$$

where R_{min} is the minimum rate requirement of each user in the system. As the T-RIS cannot amplify the signal received from the feeding antenna, the amplitude constraint of each T-RIS element is given as:

$$|\alpha_k| \leq 1, \forall k. \tag{10}$$

Then, the phase constraint of each T-RIS element is:

$$\psi_k \in [0, 2\pi], \forall k. \tag{11}$$

Furthermore, there are location constraints on the UAV, bounding it to stay within the designated service area. These constraints are given as:

$$u_x \in [u_x^{min}, u_x^{max}], \tag{12}$$

$$u_y \in [u_y^{min}, u_y^{max}], \tag{13}$$

where u_x^{min} and u_x^{max} are the minimum and maximum values of the x-coordinates of the area assigned to the UAV, and u_y^{min} and u_y^{max} are the minimum and maximum values of the y-coordinates of the area assigned to the UAV.

IV. Proposed Solution

In this section, we divide the given problem into sub-problems and systematically solve each one. Section IV-A focuses on determining the optimal beamforming at the T-RIS, Section IV-B addresses the optimal power allocation for NOMA, Section IV-C presents the framework for UAV location optimization, and Section IV-D introduces a unified optimization framework that integrates all the individual solutions.

A. Optimizing T-RIS Beamforming

To make the problem tractable, we introduce auxiliary variables H_n and F, defined as $H_n = h_n h_n^{\dagger}$ and $F = ff^{\dagger}$. Additionally, we enforce $F \geq 0$ and impose the rank(F)=1 constraint. With these auxiliary variables, the updated problem of optimizing the beamforming at the T-RIS is formulated as:

$$\max_{F} \sum_{n=1}^{N} B \log_2 \left(1 + \frac{p_n Tr(H_n F)}{Tr(H_n F) \sum_{i=1}^{n-1} p_i + \sigma^2}\right), \tag{14}$$

s.t.:

$$B \log_2 \left(1 + \frac{p_n Tr(H_n F)}{Tr(H_n F) \sum_{i=1}^{n-1} p_i + \sigma^2}\right) \geq R_{min}, \tag{15}$$

$$F \geq 0, \tag{16}$$

$$\text{Diag}(F) \leq I, \tag{17}$$

$$\text{rank}(F) = 1, \tag{18}$$

where $Tr()$ denotes the trace operator, and I is the identity matrix of order $K \times K$. The first constraint ensures the satisfaction of the minimum rate requirement for each user. The second constraint requires the matrix F to be positive semi-definite, and the third constraint specifies that no signal amplification is performed by the T-RIS elements (i.e., the amplitude at each element must be less than or equal to unity). The final constraint states that F must be a rank-one matrix, allowing us to compute the optimal value of f from the optimal F, such that $F = ff^{\dagger}$. Further simplification gives us:

$$\max_{F} \sum_{n=1}^{N} \left(B \log_2 \left(p_n Tr(H_n F) + Tr(H_n F) \sum_{i=1}^{n-1} p_i + \sigma^2\right) - B \log_2 \left(Tr(H_n F) \sum_{i=1}^{n-1} p_i + \sigma^2\right)\right), \tag{19}$$

s.t.:

$$p_n Tr(H_n F) \geq (2^{Rmin/B} - 1) \left(Tr(H_n F) \sum_{i=1}^{n-1} p_i + \sigma^2\right), \tag{20}$$

(16), (17), (18).

[1]Note that sorting at the UAV does not imply any change in the location of ground devices. This simply means that, from the UAV's perspective, device-1 is the device with the highest channel gain, device-2 has the second-highest channel gain, and so on. This sorting has no effect on the actual distribution of the users.

Although the transformation in (20) renders the constraint convex, the rank-1 constraint remains non-convex. Moreover, the objective function is the difference of concave functions. Next, we temporarily relax the rank-1 constraint during the optimization process. If the resulting solution violates the constraint, we employ a Gaussian randomization process to ensure compliance [38]. To facilitate the transformation of the objective function, we introduce auxiliary variables $\Lambda_{n,1}$ and $\Lambda_{n,2}$ as follows:

$$p_n Tr(H_n F) + Tr(H_n F) \sum_{i=1}^{n-1} p_i + \sigma^2 \geq \Lambda_{n,1}, \tag{21}$$

$$Tr(H_n F) \sum_{i=1}^{n-1} p_i + \sigma^2 \leq \Lambda_{n,2}. \tag{22}$$

Further, introducing the variables $\omega_{n,1}$ and $\omega_{n,2}$ such that: $B\log_2(\Lambda_{n,1}) \geq \omega_{n,1}$ and $B\log_2(\Lambda_{n,2}) \leq \omega_{n,2}$, the problem becomes:

$$\max_{(F,\omega_{n,1},\omega_{n,2},\Lambda_{n,1},\Lambda_{n,2})} \sum_{n=1}^{N} (\omega_{n,1} - \omega_{n,2}), \tag{23}$$

$$\text{s.t.: } B\log_2(\Lambda_{n,1}) \geq \omega_{n,1}, \tag{24}$$

$$B\log_2(\Lambda_{n,2}) \leq \omega_{n,2}, \tag{25}$$

$$(16), (17), (20), (21), (22).$$

The constraint in (25) remains non-convex. To address this, we apply a first-order Taylor approximation to transform the constraint into a standard convex form. The application of Taylor approximation results in the following expression:

$$B\log_2(\overline{\Lambda_{n,2}}) + \frac{B(\Lambda_{n,2} - \overline{\Lambda_{n,2}})}{\overline{\Lambda_{n,2}}\ln(2)} \leq \omega_{n,2}, \tag{26}$$

where $\overline{\Lambda_{n,2}}$ represents the value of $\Lambda_{n,2}$ from the previous iteration. Thus, the transformed problem is formulated as:

$$\max_{F,\omega_{n,1},\omega_{n,2},\Lambda_{n,1},\Lambda_{n,2}} \sum_{n=1}^{N} (\omega_{n,1} - \omega_{n,2}), \tag{27}$$

$$\text{s.t.: } (16), (17), (20), (21), (22), (24), (26). \tag{28}$$

The transformed problem conforms to a standard SDP format. We utilized the Mosek solver in CVXPY to solve the problem within the Python environment using the interior-point method.

Discrete Phase-Shift: Although the framework proposed in the previous section provides an optimal solution for beamforming at the T-RIS, for large T-RIS with a significant number of elements, it becomes more practical to implement only discrete phase shifts at the RIS [36]. Note that this discretization of the phase shifts introduces sub-optimality and can cause performance degradation. However, for very large T-RIS, adjusting continuous phase-shift values for each element may become impractical.

Hence, after obtaining the optimal beamforming solution from the framework proposed in the previous section, the discretization of the phase shifts can be performed by rounding the continuous phase-shift values to the closest discrete value. The set of discrete values is given as:

$$\mathcal{S} = \{0, \Delta\psi, 2\Delta\psi, \ldots, (Z-1)\Delta\psi\},$$

where Z denotes the number of phase shift levels and $\Delta\psi = \frac{2\pi}{Z}$.

Computational Complexity: The computational complexity of solving the problem using the Mosek solver in CVXPY is $O(DNK^{3.5})$, where D is the number of iterations required for convergence [37].

B. Optimizing Power Allocation

The power allocation problem for the considered system model is written as:

$$\max_{p_n} \sum_{n=1}^{N} B\log_2\left(1 + \frac{p_n |h_n^\dagger f|^2}{|h_n^\dagger f|^2 \sum_{i=1}^{n-1} p_i + \sigma^2}\right), \tag{29}$$

$$\text{s.t. } (8), (9).$$

For $N = 2$, the problem is easily solvable and has been addressed multiple times in the literature using different approaches. However, as the number of users increases, finding a closed-form expression for the solution becomes particularly complex. The challenge arises due to the interdependence of power allocation values among devices, leading to an exponential increase in the order of the closed-form expression's complexity with the number of users in the system. Since we aim to solve the problem for any N, where N can be greater than 2, we introduce a variable Q_n such that $Q_n = \sum_{i=1}^{n} p_n$. With this introduction, the objective function can be transformed as:

$$\sum_{n=1}^{N} B\log_2\left(1 + \frac{p_n |h_n^\dagger f|^2}{|h_n^\dagger f|^2 \sum_{i=1}^{n-1} p_i + \sigma^2}\right) = \sum_{n=1}^{N} B\log_2\left(\frac{|h_n^\dagger f|^2 \sum_{i=1}^{n} p_i + \sigma^2}{|h_n^\dagger f|^2 \sum_{i=1}^{n-1} p_i + \sigma^2}\right). \tag{30}$$

Substituting $Q_n = \sum_{i=1}^{n} p_n$ gives us:

$$\sum_{n=1}^{N} B\log_2\left(\frac{Q_n |h_n^\dagger f|^2 + \sigma^2}{Q_{n-1} |h_n^\dagger f|^2 + \sigma^2}\right), \tag{31}$$

rearranging (31) we get:

$$\sum_{n=1}^{N} B\log_2\left(\frac{Q_n |h_n^\dagger f|^2 + \sigma^2}{Q_n |h_{n+1}^\dagger f|^2 + \sigma^2}\right). \tag{32}$$

With this transformation the problem can then be written as:

$$\max_{Q_n} \sum_{n=1}^{N} B\log_2\left(\frac{Q_n |h_n^\dagger f|^2 + \sigma^2}{Q_n |h_{n+1}^\dagger f|^2 + \sigma^2}\right), \tag{33}$$

$$\text{s.t.: } Q_n |h_n^\dagger f|^2 + \sigma^2 \geq 2^{R_{min}/B}(Q_{n-1} |h_n^\dagger f|^2 + \sigma^2), \forall n \tag{34}$$

$$Q_N \leq P_{max}, \tag{35}$$

$$Q_n \leq Q_{n+1}, \forall n = 1, 2, 3 \ldots, N-1, \tag{36}$$

where the first constraint ensures the satisfaction of the rate requirement for each device, the second constraint guarantees

adherence to the power budget for UAV, and the last constraint ensures non-negativity of the power allocated to each user. Since the transformed problem is convex, we employ dual decomposition to compute the optimal value of Q_n. Substituting $g_n = |h_n^\dagger f|^2$, the Lagrangian of the problem is expressed as:

$$\begin{aligned} L = &- \sum_{n=1}^{N} \log_2\left(\frac{Q_n g_n + \sigma^2}{g_{n+1} Q_n + \sigma^2}\right) + \eta(Q_N - P_{max}) \\ &+ \sum_{n=1}^{N-1} \lambda_n (Q_n - Q_{n+1}) \\ &+ \sum_{n=1}^{N} \tau_n \left(2^{R_{min}/B}(g_n Q_{n-1} + \sigma^2) - Q_n g_n - \sigma^2\right). \end{aligned} \tag{37}$$

Applying KKT conditions [39], we compute the closed-form expression for the optimal Q_n as:

$$Q_n^* = \begin{cases} \dfrac{X_n \pm \sqrt{Y_n}}{Z_n} & \text{if } n = 1, ..N-1, \\ \dfrac{g_n B + (-\lambda_{n-1} + \eta)\sigma^2 - g_n \sigma^2 \tau_n}{g_n(\lambda_{n-1} - \eta + g_n \tau_n)} & \text{otherwise,} \end{cases} \tag{38}$$

where $X_n = (g_n + g_{n+1})\sigma^2(\lambda_n - \lambda_{n-1} - g_n \tau_n + 2^{R_{min}/B} g_{n+1}\tau_{n+1})$, $Y_n = \sigma\sqrt{g_n - g_{n+1}}\sqrt{-\lambda_n + \lambda_{n-1} + g_n\tau_n - 2^{R_{min}/B} g_{n+1}\tau_{n+1}}\sqrt{\zeta_n}$, $Z_n = 2g_n g_{n+1}(-\lambda_n + \lambda_{n-1} + g_n\tau_n - 2^{R_{min}/B} g_{n+1}\tau_{n+1})$, for $\zeta_n = g_n^2\sigma^2\tau_n + g_{n+1}\sigma^2(\lambda_n - \lambda_{n-1} + 2^{R_{min}} g_{n+1}\tau_{n+1}) - g_n((\lambda_n - \lambda_{n-1})\sigma^2 + g_{n+1}(-4B + \sigma^2\tau_n + 2^{R_{min}/B}\sigma^2\tau_{n+1}))$. The proposed transformation offers an advantage in that the order of the closed-form expression in (38) does not increase with N unlike the solutions proposed in the literature for multiuser NOMA power allocation [40], [41]. We then employ the sub-gradient method to optimize the values of the dual variables, where, in each iteration, the values are updated as:

$$\eta(t+1) = \eta(t) + \delta(Q_N - P_{max}), \tag{39}$$

$$\lambda_n(t+1) = \lambda_n(t) + \delta(Q_n - Q_{n+1}), \tag{40}$$

$$\tau_n(t+1) = \tau_n(t) + \delta(2^{R_{min}/B}(g_n Q_{n-1} + \sigma^2) - Q_n g_n - \sigma^2), \tag{41}$$

where t denotes the iteration number, and δ is the step size or learning rate of the subgradient method. Then, from the optimal values of Q_n, the optimal value of p_n is computed as:

$$p_n^* = Q_n^* - Q_{n-1}^*. \tag{42}$$

Computational Complexity: The computational complexity of solving the problem using the proposed Lagrangian dual technique is $O(3EN)$, where E is the number of iterations until the proposed framework converges.

C. Optimizing the Location of UAV

For a predetermined height of the UAV, the problem of optimizing the location of the UAV subject to rate and coordinate constraints is formulated as:

$$\max_{u_x, u_y} \sum_{n=1}^{N} B \log_2\left(1 + \frac{p_n |h_n^\dagger f|^2}{|h_n^\dagger f|^2 \sum_{i=1}^{n-1} p_i + \sigma^2}\right), \tag{43}$$

s.t.: (9), (12), (13).

The considered problem is strictly non-convex in the coordinates of the UAV, as evident from (1), (4), (5). Consequently, the schemes used earlier in the paper are not applicable to solve this problem. In this section, we propose a dueling-based double DRL scheme (D3RL) to find the coordinates that maximize the sum rate of the system while satisfying all system requirements. Leveraging the fact that the UAV, equipped with a T-RIS, can sense the location of the ground devices [42], [43], we train an RL agent using the ground devices' locations to optimize the UAV's location.

In the considered system, the sum rate depends not only on controllable parameters but also on the locations and the number of the devices and their distributions. For instance, a system where ground devices are located near each other offers a higher sum rate due to better channels compared to a system where devices are located far apart. In such systems, using a simple DRL for optimization becomes complicated, as the agent needs to understand that different locations of the ground devices would impact the sum rate of the system at the optimal value of optimization variables. In problems like these, the dueling mechanism significantly boosts the learning process by separating the value associated with the positions of the ground devices from the advantage of moving the UAV to a particular location. The function associated with the current state (locations of the ground devices) of the system is called the value function, whereas the function associated with the advantage of taking a particular action (moving UAV) is called the advantage function in dueling-based DRL.

In DRL frameworks, three important sets/values are part of the training process: the state set, the action set, and the reward. The state set is provided as input to the agent and contains essential information about the current state of the system (locations of the ground devices in our case). The action set encompasses all possible actions that the agent can take (e.g., moving in a particular direction in our case). Finally, the reward is the value that serves as feedback to the agent, aiding in learning what constitutes a good action in a given state (in our case, this is the difference in sum rates of the system before and after taking the action, considering all constraints are satisfied).

In DRL frameworks, the Q-value function is commonly employed to estimate the cumulative reward of taking an action, comprising the current reward and potential future rewards resulting from that action. Deep neural networks (DNNs) are utilized in DRL to predict the Q-value associated with taking a particular action, and the action with the maximum Q-value is selected as the optimal action. This mapping from the state to the optimal action, performed by the DRL agent, is referred to as the agent's policy. In double deep RL, separate DNNs are employed to estimate the Q-values for actions and the future rewards of taking an action; these DNNs are termed the *primary DNN* and *target DNN*, respectively. This separation mitigates the problem of overestimation in DRLs, contributing to a more stable training framework. The

Q-value computation is then expressed as:

$$Q(s,a) = \Omega(s,a) + \xi \max_{\overline{a}\sim\overline{\pi}} Q(\overline{s},\overline{a}). \tag{44}$$

In the Q-value expression (44), the term $\Omega(s,a)$ represents the immediate reward of taking action a while the system is in state s. The second term, known as discounted future rewards, incorporates the impact of future rewards. Here, ξ denotes the discount factor, with smaller values of ξ implying reduced influence of future rewards on training, while larger values increase their impact. The expression $\max_{\overline{a}\sim\overline{\pi}} Q(\overline{s},\overline{a})$ within the discounted reward term is provided by the *target DNN*. It offers an estimate of the future Q-values or rewards when optimal actions are taken based on the current policy of the *target DNN*. Here, $\overline{a}$ represents future actions, and $\overline{\pi}$ is the policy of the *target DNN*.

In the proposed D3RL framework, as defined earlier, the value function and advantage function are computed as follows:

$$V^{\pi}(s) = \mathop{\mathbb{E}}_{a\sim\pi}\{Q^{\pi}(s,a)\}, \tag{45}$$

$$A^{\pi}(s,a) = Q^{\pi}(s,a) - V^{\pi}(s). \tag{46}$$

With the value function and advantage function, the Q-value is computed as:

$$Q(s,a,\Pi) = V(s,\Pi) + \left(A(s,a,\Pi) - \frac{1}{\mathcal{X}}\sum_{a} A(s,a,\Pi)\right), \tag{47}$$

where $\mathcal{X}$ denotes the size of action set, and Π denotes the weights and biases of the *primary DNN*. The inclusion of Π in Eq. (47) shows that all the values used are obtained from the *primary DNN*.

Then, the *primary DNN* is trained to minimize the difference between the values obtained from (44) and (47). This results in training the D3RL agent to provide accurate estimates of the Q-values associated with taking any action in a given state.

The details of the reward, state, and action sets used in the proposed technique are as follows:

- *State:* The area allocated to the UAV is partitioned into small grids, forming a grey-scale image provided to the agent. Each grid is represented as a pixel, and the pixel brightness reflects the number of ground devices in that grid. The default pixel value is set to 0 (black). For each ground device in the area, 0.1 is added to the pixel value, enhancing brightness with an increase in the number of users in that location. An example of the system state provided to the agent is illustrated in Fig. 2. Additionally, each action performed by the agent influences the state. For instance, if the agent moves one block up in the y-direction, all bright pixels will shift one block down. To accommodate a service area of W rows and W columns allocated to the UAV, $W/2$ rows of black pixels are appended at the top and bottom of the state figure, and $W/2$ columns of black pixels are added to the left and right, ensuring a complete state representation even when the UAV is at the boundary of the allocated area.

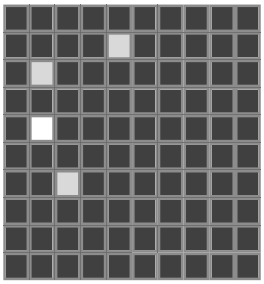

Fig. 2. The illustration depicts a snapshot of the system's state. In the grid, black squares signify the absence of devices, while lighter blocks indicate the presence of ground devices in the corresponding areas. Notably, the square located in the fifth row and second column appears brighter than other blocks, indicating a higher concentration of devices in this specific area compared to others.

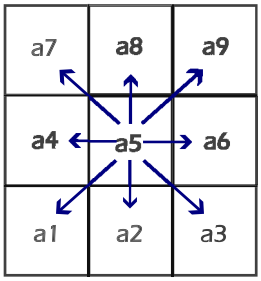

Fig. 3. This figure offers a comprehensive overview of the potential actions available to the D3RL agent. In this context, action $a1$ signifies the UAV's movement to the lower-left block in the grid, $a2$ denotes the UAV's shift to the lower block, selecting $a3$ prompts the UAV to move to the lower-right block, opting for action $a4$ directs the UAV to the block on the left, action $a5$ signifies the UAV's decision to remain stationary in the current block, action $a6$ leads the UAV to the block on the right, the selection of action $a7$ results in the UAV moving to the upper-left block, action $a8$ entails the UAV's relocation to the block directly above the current location, and finally, action $a9$ indicates the UAV's movement to the upper-right block.

- *Actions:* The proposed solution scheme allows the agent to take any of the 9 actions denoted as $\{a1, a2, a3, \ldots, a9\}$. Each action corresponds to a specific movement, such as moving to the lower-left block ($a1$), moving to the lower block ($a2$), moving to the lower right block ($a3$), moving to the block on the left ($a4$), deciding to remain in the same block ($a5$), and so on. Fig. 3 provides a comprehensive visualization of each action.
- *Reward:* At the conclusion of each action, the agent receives a reward as feedback to aid in learning the effectiveness of the action. In the proposed framework, the reward is computed at the current location after taking the action by solving the optimization problems related to beamforming and power allocation, as explained in the preceding subsections. If a feasible solution exists at the current location, the reward is determined by the difference in the sum rate of the system at the current location after the action and the sum rate at the previous location. This guides the agent to move to adjacent blocks that increase the system's sum rate. If there is no adjacent block offering a higher sum rate, the agent decides to stay idle, with the reward for staying in the same block set to 0. Fig. 4 illustrates examples of rewards the agent

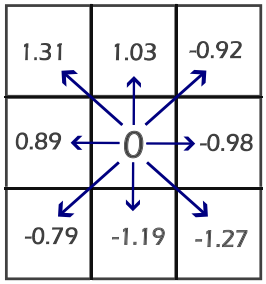

Fig. 4. The figure illustrates an example of rewards provided to the D3RL agent as a result of taking each action. It showcases that the least favorable action for the UAV is moving to the lower-right block, resulting in the minimum reward. In contrast, the optimal action, yielding the maximum reward, involves moving to the upper-left block. Furthermore, the figure demonstrates that remaining stationary in the current block results in an anticipated reward value of 0.

might receive for different actions. If a feasible solution does not exist or a constraint is violated, the reward value is set to a negative number.

1) DNN Model: The system's state is provided as a grey-scale image, necessitating the use of a convolutional layer with 20 kernels, each sized 3×3, as the first layer in the proposed DNN. Following this convolutional layer is the flattening layer, responsible for converting the 2-dimensional outputs of the convolutional kernels into a stack of 1-dimensional arrays. Subsequently, two dense hidden layers, sized 1000 nodes and 500 nodes, follow the flattening layer. The output layer, appearing after the second hidden layer, contains 10 nodes. Specifically, 9 of these nodes provide the advantage values for each action $(a1, a2, \ldots, a9)$, while the 10th node offers the value associated with the state. These outputs are then fed into the dueling mechanism, executing the operations detailed in (47), ultimately providing the estimate of the Q-value associated with each action in the action space. The steps and the proposed DNN model are visually represented in Fig. 5. Further, rectified linear-unit (ReLu) activation function is applied in the nodes of all layers, except for the output layer where the linear activation function is used.

2) D3RL Training: The steps involved in training the D3RL framework are provided in Algorithm 1. We propose a replay memory-based training framework where the agent saves the state, action taken, the reward obtained, and the next state after taking the action, in memory. In each training round, we sample a fixed-size mini-batch from the memory, and the agent is trained on the mini-batch to minimize the Q-value estimation error. The steps in the algorithm are explained below.

The training of the D3RL agent depends on exploring new actions to update the policy in each training round. In the proposed algorithm, we introduce a soft exploration technique where actions are drawn in relation to the current policy depending on the current temperature of exploration. At small values of exploration temperature, the probability of taking any action is almost the same. With each iteration, the temperature of exploration is increased, resulting in the agent taking actions that have a higher probability of returning higher rewards.

Algorithm 1: Training the D3QL Agent

1) Initialize the memory and set υ=0
2) Initialize *primary DNN* with weights Π and *target DNN* with $\overline{\Pi}$
3) **for** each step
4) Observe state (s)
5) Take action a by doing *soft exploration*
6) Increase the value of υ as $\upsilon = \upsilon + \varpi$
7) Receive feedback and compute the reward
8) Store a, s, $\bar{s}$, and reward in the memory.
9) Draw I samples from the memory for training the *primary DNN*
10) For each sample compute $Q(s, a)$ using the value of obtained reward and employ *target DNN* for the estimated future reward as in (44)
11) Compute agent's estimations of Q-values using the *primary DNN* as in (47), then calculate the loss and perform back-propagation on the *primary DNN* to minimize the estimation error
12) Update $\overline{\lambda}$ as $\overline{\Pi} = (\Pi + \overline{\Pi})/2$
13) **end for**

In soft exploration, the probability distribution of actions is computed as:

$$Pa = \frac{e^{Q_{a_k} \times \upsilon}}{\sum_{k=1}^{\chi} e^{Q_{a_k} \times \upsilon}}$$

where Q_{a_k} is the expected Q-value of taking action a_k in the current state, and υ denotes the temperature of exploration. Then, during soft exploration, the actions are drawn according to $a \sim Pa$. With each training round, the value of υ is increased, resulting in reducing the probability of random actions.

In the first step of the algorithm, we initialize the memory buffer and set the value of the soft exploration temperature $\upsilon = 0$. Then, we initialize the *primary DNN* and *target DNN* with random weights. In each training step, the agent is provided the current state of the system in Step 4. In Step 5, the agent performs soft exploration, and the exploration temperature is incremented by ϖ in Step 6. Then, the agent receives the reward in Step 7, which is computed as the difference in the sum rate of the system before and after taking the action, as explained earlier. The state, action, reward and next state are stored in the memory buffer in Step 8. In Step 9, the agent randomly selects I training samples from the memory buffer. In Step 10, the agent computes the Q-values for each sample using *target DNN* and the reward value as shown in Equation (44). Then, we compute the estimated Q-values of the action using *primary DNN* as shown in Equation (47) and perform back-propagation to minimize the estimation error of the *primary DNN*. In Step 12, the weights of *target DNN* are set equal to the average of the current weights of the *target DNN* and the updated weights of the *primary DNN*.

Computational Complexity: For the service area of the UAV divided into grids, the computational complexity of training

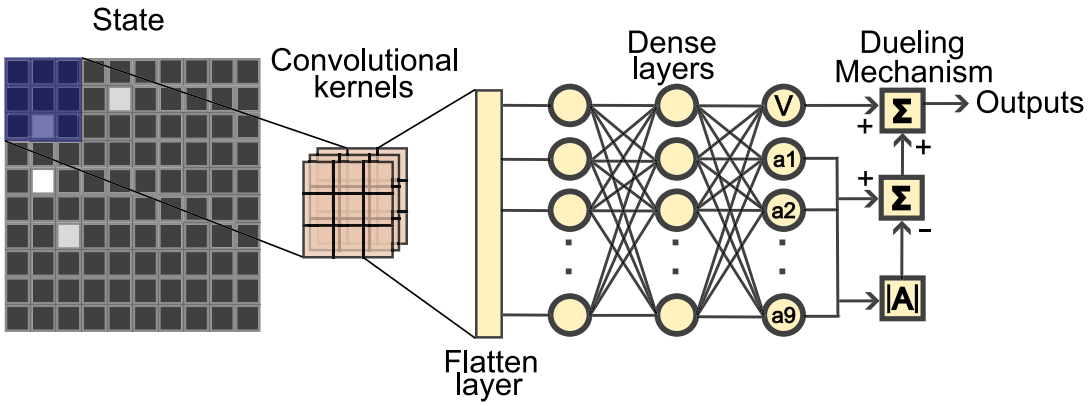

Fig. 5. The figure demonstrates the application of the DNN structure to the system's state and illustrates how the dueling mechanism is employed to estimate the Q-value of each action.

the proposed D3RL agent is given as:

$$O\Big(T_e\Big(V(W-(L-1))^2 \times A + A \times B + B \times 10\Big)\Big)$$

where T_e is the number of training rounds or epochs required to train the agent, V denotes the number of kernels in the convolutional layer, L is the number of rows and columns in each convolutional kernel, W is the dimension of the input state (i.e., the number of pixels in each row and in each column), A denotes the number of nodes in the first dense layer, B is the number of nodes in the second hidden layer, and there are 10 nodes in the output layer [44].

D. Joint Solution Scheme

After the training of the D3RL agent is finished, the trained model can be used to optimize the location of the UAV, along with the solutions provided in previous sections to optimize the beamforming and power allocation. In this section, we provide Algorithm 2 that combines all the schemes to provide a joint optimization.

In the algorithm, we have nested loops. The outer loop keeps the algorithm running until convergence, meaning that the algorithm runs until the change in the average rate of two consecutive iterations is very small. In Step 2, we provide the UAV with the current state of the system (locations of the ground devices). Then, in Step 3, we update the location of the UAV according to the output of the D3RL agent. In Step 4, we get the channels to the ground devices at the current location and generate a random beamforming vector, which is then passed to the inner loop of Step 5.

The inner loop represents an alternate optimization framework, which runs until convergence or until we reach a maximum number of iterations. In the inner loop, in Step 6, we use the beamforming value to compute the optimal power allocation using the scheme proposed in Section IV-B. In Step 7, we use the optimal power allocation value obtained from Step 6 to optimize beamforming. This beamforming solution is then provided to the optimization in Step 6 to optimize the power allocation according to the updated beamforming vector. The alternate optimization in Step 6 and Step 7 continues until we reach the maximum number of iterations or until the convergence requirements are satisfied, as shown in Step 8. Then, we check for the convergence of the outer loop in Step 9. At the convergence of the outer loop, the solution values are returned in Step 10.

Algorithm 2: Joint Optimization Algorithm

1) **Until** convergence do
2) Provide current state of the system to the trained agent
3) Update location of the UAV according to the output of the D3RL agent trained as in Section IV-C
4) Get values of h_n at the current location and generate a random beamforming vector
5) **for** S iterations OR until convergence do
6) Optimize power allocation for the given beamforming vector as proposed in Section IV-B
7) For the power allocation values obtained from previous step, optimize the beamforming vector using the technique proposed in Section IV-A
8) If convergence requirements of inner loop are satisfied break
9) Check for convergence requirements of the outer loop
10) *Return* optimal beamforming value, optimal power allocation and optimal location of the UAV

Computational Complexity of the Joint Framework: The computational complexity of the joint framework is: $O(I_1(V(W-(L-1))^2 \times A + A \times B + B \times 10 + I_2(DNK^{3.5} + 3EN)))$, where I_1 is the number of iterations required for the convergence of the outer loop of Algorithm 2, I_2 is the number of iterations required for the convergence of the inner loop. V is the number of kernels in the convolutional layer, L is the number of rows and columns in each convolutional kernel, W is the dimension of the input state (i.e., number of pixels in each row and in each column), A and B denote the number of nodes in the first and second dense layers, respectively. Then, D and E are the numbers of iterations

TABLE I
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
ϱ	3	σ^2	0.001
f_c	3.4 GHz	R_{min}	10 kbits/s
B	1 GHz	P_{max}	10 W
N	4	u^z	20 m
M	9	K_n	5
Ω	$\frac{c}{10f_c}$		

required for the techniques in Sections IV-A and IV-B to converge, respectively.

V. SIMULATION RESULTS AND DISCUSSION

For the simulations, we consider that the UAV is allocated an area of 10,000 m^2 ($100m \times 100m$). We divide the area into small grids of dimensions $5m \times 5m$, resulting in a grid of 20 rows and 20 columns. The values of the remaining system parameters are provided in Table I, unless stated otherwise. Considering a practical scenario, we assume that the distribution of the ground devices follows a hotspot area, which can be modeled as a two-dimensional truncated Gaussian distribution [45], where the center point is selected at random in the allocated area, and the variance of the distribution was taken to be 25.

To highlight the benefits of location optimization, we compare the performance of the proposed joint optimization scheme with a benchmark where the UAV is located at the center of the allocated area, and the location of the UAV is not optimized. However, in the benchmark, the beamforming and power allocation at the UAV are optimized as proposed in Sections IV-A and IV-B. The fixed location scheme is referred to as LOC Fix in the simulation section, and the proposed joint optimization scheme is referred to as LOC Opt. The D3RL agent was trained for 10,000 samples, and all the results presented in this section are averaged over 1000 samples.

In NOMA transmissions, the sum rate of the system is largely dependent on the channel gain of the closest ground device (having the best channel conditions), as most of the system's resources are allocated to the device with the best channel conditions, such that all the constraints are satisfied. Further, as the number of devices in the system increases, the interference temperature of the system also increases, which negatively impacts the sum rate. Fig. 6 shows the impact of an increasing number of ground devices in the system on the sum rate. For the proposed LOC Opt framework, when the number of devices is increased, the sum rate of the system decreases as more power is allocated to satisfy the rate requirements of the weak users (users with relatively bad channel conditions), leaving behind less power available for the transmission to the closest user (the closest user with the best channel conditions also becomes more significant because it performs SIC to remove interference from the signals of all other devices in the system). Hence, increasing the number of users decreases the sum rate of the system.

However, it is interesting to see that for the case where the position of the UAV is fixed in the center of the allocated area, as we increase the number of ground devices, the sum rate also increases. The reason behind the increase in sum rate is

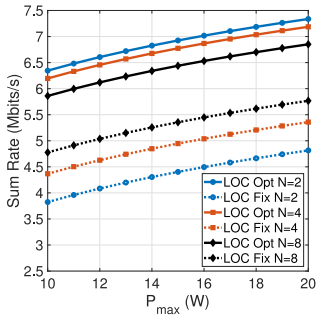

Fig. 6. System's performance for increasing value of P_{max} under varying the number of ground devices following a hotspot area distribution.

that, for the fixed location of the UAV, when we have more users in the system, the probability that the distance of the closest user from the UAV will be small also increases. For example, for 8 ground devices in the system, there is a high probability that a device will be close to the UAV compared to the system with just 2 ground devices. As a result, the system allocates most of its resources to the closest users while satisfying all other constraints of the system, and the sum rate of the system increases. However, it can be seen that for the LOC Fix scheme, the marginal gain of the system (marginal gain is the unit gain in the sum rate when a new ground device is added) decreases with an increasing number of ground devices because, for a large number of ground devices, the disadvantage of allocating more power to the weak users starts to take over the advantage of the devices being located closer to the UAV. Fig. 6 also shows that increasing the value of $P_{\max}$ also increases the sum rate of the system as more power becomes available for transmission. Further, for any number of ground devices in the system, the proposed LOC Opt framework outperforms the LOC Fix scheme.

All the results in this section consider a hotspot area of the ground devices, which is modeled as a truncated Gaussian distribution. However, for a complete analysis, in Fig. 7, we show the impact of an increasing number of users and $P_{\max}$ on the sum rate of the system when the location of ground devices follows a uniform distribution. It can be seen that, just like the results in Fig. 6, for the LOC Opt scheme, in the case of a uniform distribution, the sum rate of the system decreases with increasing N due to a rise in the interference temperature of the system. On the other hand, similar to the hotspot distribution, the sum rate of the system for a uniform distribution of ground devices increases in the case of LOC Fix as the probability that the strongest user (user closest to the UAV) will have a better channel also increases. However, the marginal gain in the sum rate in the case of LOC Fix decreases with increasing users because the disadvantage of increasing interference temperature starts to shadow the advantage due to the better channel of the strongest user.

Another interesting thing to note is that the gap in the sum rate offered by LOC Opt and LOC Fix for $N = 8$ in the case of a uniform distribution is less as compared to the hotspot distribution, which also makes sense intuitively. For a uniform distribution where a large number of ground devices are scattered all over the area, the optimal location of the UAV

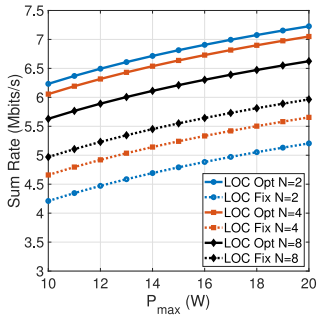

Fig. 7. System's performance for increasing value of P_{max} under varying the number of ground devices following a uniform distribution.

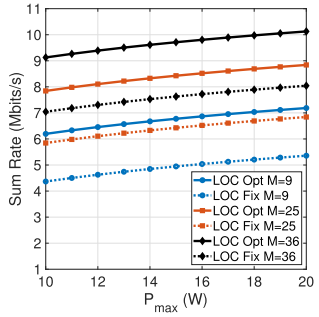

Fig. 8. System performance for increasing P_{max} under different M.

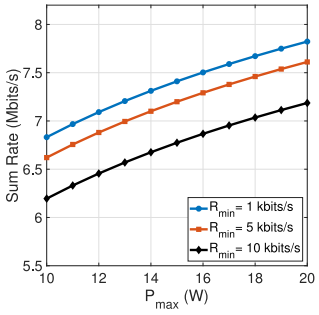

Fig. 9. System's sum-rate performance under different values of R_{min}.

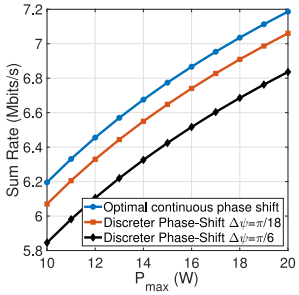

Fig. 10. The effect of phase-shift discretization at the T-RIS.

to serve all the devices would be the center of the service area. However, for a practical number of users being served on the same channel in NOMA, it can be seen that LOC Opt always outperforms LOC Fix, for any value of $P_{\max}$.

The impact of increasing the number of T-RIS elements on the sum rate is shown in Fig. 8. As we increase the number of T-RIS elements mounted on the UAV, the sum rate of the system also increases. However, it should be kept in mind that, as we are assuming an T-RIS-mounted UAV, it might not be possible to have a very large number of T-RIS elements on the UAV, and factors like the load-bearing capacity of the UAV should be accounted for while deciding the size of the T-RIS. Further, it can be seen that for $M = 36$, LOC Fix provides more rate than LOC Opt with 9 T-RIS elements. However, for the same number of T-RIS elements, the proposed LOC Opt framework always outperforms the LOC Fix scheme.

The impact of increasing $R_{\min}$ on the sum rate of the system for different values of $P_{\max}$ is shown in Fig. 9. When the minimum rate requirement of the ground devices is increased, more power is allocated for the signals of the weak users, leaving behind less power for the signal of the strong users, resulting in a decrease in the sum rate of the system. Further, as the rate of a user is a logarithmic function of the allocated power, a linear increase in $R_{\min}$ results in an exponential increase in the required power to satisfy the rate requirement of the device. Hence, it can be seen that the decrease in the sum rate when $R_{\min}$ is increased from 1 kbits/s to 5 kbits/s is much less compared to when $R_{\min}$ is increased from 5 kbits/s to 10 kbits/s. As increasing the value of $R_{\min}$ bounds the system to allocate more resources to the weak user, the system with a smaller value of $R_{\min}$ will always provide higher values of the sum rate as compared to a system with higher values of $R_{\min}$.

As discussed earlier in Section IV-A, in some systems with limited resources, it may not be feasible to employ the continuous phase-shift-based optimal solution. Hence, discretization of the phase shifts is required, where each element may be restricted to discrete levels of phase shifts. The Fig. 10 shows the impact of discretization of the phase shifts on the sum rate of the system. As expected, discretization introduces sub-optimality, resulting in performance degradation. When discretization is introduced with each discrete step incremented by a value of $\frac{\pi}{18}$, the sum rate of the system decreases compared to the optimal continuous phase-shift values. Further increasing the discretization step from $\Delta\psi = \frac{\pi}{18}$ to $\Delta\psi = \frac{\pi}{6}$ causes a further decrease in the sum rate. However, it should be noted that the performance degradation due to phase-shift discretization at the T-RIS results from system limitations and does not imply any sub-optimality in the proposed solution.

Although the considered problem has not been addressed in the literature before, some works have proposed deep Q-learning (DQL)-based schemes for NOMA power allocation [46], [47]. In DQL-based NOMA power allocation, system parameters such as the users' channel state information are provided to the DNN as the system state. The action space includes actions like increasing or decreasing the power allocated for each user's transmission. The reward, in this case, is the sum rate of the system, provided that no constraints are violated; otherwise, the reward is zero. The Fig. 11 presents a comparison of the proposed framework with a DQL-based scheme for NOMA power allocation, where T-RIS beamforming and UAV positioning are optimized as proposed in this work. Since our proposed framework yields optimal power allocation values, it outperforms the DQL framework

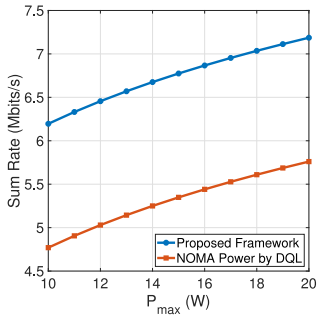

Fig. 11. Comparison of the proposed framework with the technique where NOMA power allocation is optimized using DQL.

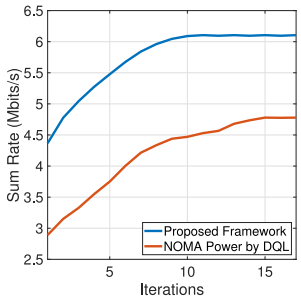

Fig. 12. Convergence behavior of the proposed scheme and the technique where NOMA power is allocated using the DQL.

across all values of P_{max}, highlighting the benefits of the proposed NOMA power allocation scheme over DQL-based approaches.

Fig. 12 shows the convergence of the joint solution framework, proposed in Section IV-D, and the scheme with DQL-based NOMA power allocation. At the start of the optimization process, the UAV is positioned at the center of the allocated area. In each iteration, the proposed framework updates the UAV's location and calculates the power allocation values and beamforming vector. As shown in Fig. 12, the sum rate of the system increases with each iteration for both schemes, demonstrating their efficiency. The sum rate continues to rise until it reaches a maximum value, where the solution converges. Additionally, it is clear that at convergence, the proposed framework achieves a higher sum rate than the DQL-based scheme.

VI. Conclusion

In this study, we proposed a joint optimization framework for optimizing the location of a T-RIS-mounted NOMA-enabled UAV system, the T-RIS beamforming vector, and the power allocation for NOMA transmission. The optimization process took into account the rate requirements for each ground device within the UAV's designated area, the UAV's power budget, location constraints, and practical limitations associated with T-RIS beamforming. The simulation results showed that the proposed framework provides efficient performance. We presented the results while considering both hotspot areas and the uniform distribution of the ground devices. Furthermore, optimizing UAV location was found to be critical, as the results show that optimizing the location provides a gain of up to 65.9% in the sum rate of the system. In the future, we aim to optimize channel allocation and user clustering at the UAV to limit the number of ground devices assigned to the same channel, as increasing this number raises system interference. Additionally, for a multi-UAV system, where each UAV serves its dedicated area, we plan to use federated learning to accelerate the training process of the D3RL framework.

References

[1] M. Matracia, M. A. Kishk, and M.-S. Alouini, "UAV-aided post-disaster cellular networks: A novel stochastic geometry approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9406–9418, Jul. 2023.

[2] Y. B. Jung, S. Y. Eom, and S. I. Jeon, "Experimental Design of mobile satellite antenna system for commercial use," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 429–435, May 2010.

[3] Z. Ali, Z. Rezki, and M.-S. Alouini, "Optimizing power allocation in HAPs assisted leo satellite communications," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 2, pp. 1661–1677, Nov. 2024, doi: 10.1109/TMLCN.2024.3491054.

[4] X. Huang, Y. Zhang, Y. Qi, C. Huang, and M. S. Hossain, "Energy-efficient UAV scheduling and probabilistic task offloading for digital twin-empowered consumer electronics industry," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2145–2154, Feb. 2024.

[5] P. Du, Y. Shi, H. Cao, S. Garg, M. Alrashoud, and P. K. Shukla, "AI-enabled trajectory optimization of logistics UAVs with wind impacts in smart cities," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3885–3897, Feb. 2024.

[6] M. Alzenad, A. El Keyi, F. Lagum, and H. Yanikomeroglu, "3-D placement of an unmanned aerial vehicle base station (UAV-BS) for energy-efficient maximal coverage," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 434–437, Aug. 2017.

[7] T. V. Nguyen, H. D. Le, and A. T. Pham, "On the design of RIS–UAV relay-assisted hybrid fso/rf satellite–aerial–ground integrated network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 757–771, Apr. 2023.

[8] W. U. Khan, E. Lagunas, A. Mahmood, S. Chatzinotas, and B. Ottersten, "RIS-assisted energy-efficient leo satellite communications with NOMA," *IEEE Trans. Green Commun. Netw.*, vol. 8, no. 2, pp. 780–790, Jun. 2024.

[9] W. U. Khan, A. Mahmood, C. K. Sheemar, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for 6G non-terrestrial networks: Assisting connectivity from the sky," *IEEE Internet Things Mag.*, vol. 7, no. 1, pp. 34–39, Jan. 2024.

[10] X. Xi, X. Cao, P. Yang, J. Chen, T. Quek, and D. Wu, "Joint user association and UAV location optimization for UAV-aided communications," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1688–1691, Dec. 2019.

[11] C. Pan, H. Ren, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint blocklength and location optimization for URLLC-enabled UAV relay systems," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 498–501, Mar. 2019.

[12] R. Chen, Y. Sun, L. Liang, and W. Cheng, "Joint power allocation and placement scheme for UAV-assisted IoT with QoS guarantee," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 1066–1071, Jan. 2022.

[13] C. Qiu, Z. Wei, Z. Feng, and P. Zhang, "Joint resource allocation, placement and user association of multiple UAV-mounted base stations with in-band wireless backhaul," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1575–1578, Dec. 2019.

[14] M. Asif, A. Ihsan, W. U. Khan, A. Ranjha, S. Zhang, and S. X. Wu, "Energy-efficient beamforming and resource optimization for AmBSC-assisted cooperative NOMA IoT networks," *IEEE Internet Things J.*, vol. 10, no. 14, pp. 12434–12448, Jul. 2023.

[15] L. Bing, Y. Gu, L. Hu, Y. Yin, and J. Wang, "MIMO-NOMA-aided healthcare IoT networking: Automated massive connectivity protocol," *IEEE Trans. Consum. Electron.*, vol. 69, no. 4, pp. 697–708, Nov. 2023.

[16] H. Zhang, J. Zhang, and K. Long, "Energy efficiency optimization for NOMA UAV network with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2798–2809, Dec. 2020.

[17] R. Zhang, X. Pang, J. Tang, Y. Chen, N. Zhao, and X. Wang, "Joint location and transmit power optimization for NOMA-UAV networks via updating decoding order," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 136–140, Jan. 2021.

[18] X. Liu et al., "Placement and power allocation for NOMA-UAV networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 965–968, Jun. 2019.
[19] K. Singh, P. C. Wang, S. Biswas, S. K. Singh, S. Mumtaz, and C.-P. Li, "Joint active and passive beamforming design for RIS-aided IBFD IoT communications: QoS and power efficiency considerations," *IEEE Trans. Consum. Electron.*, vol. 69, no. 2, pp. 170–182, May 2023.
[20] S. Chen et al., "Optimal RIS allocations for PLS with uncertain jammer and eavesdropper," *IEEE Trans. Consum. Electron.*, vol. 69, no. 4, pp. 927–936, Nov. 2023.
[21] M. Asif, X. Bao, A. Ihsan, W. U. Khan, M. Ahmed, and X. Li, "Securing NOMA 6G communications leveraging intelligent omni-surfaces under residual hardware impairments," *IEEE Internet Things J.*, vol. 11, no. 14, pp. 25326–25336, Jul. 2024.
[22] G. Chen, Q. Wu, W. Chen, D. W. K. Ng, and L. Hanzo, "IRS-aided wireless powered MEC systems: TDMA or NOMA for computation offloading?" *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1201–1218, Feb. 2023.
[23] S. Zeng et al., "Reconfigurable intelligent surfaces in 6G: Reflective, transmissive, or both?" *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2063–2067, Jun. 2021.
[24] W. U. Khan, E. Lagunas, A. Mahmood, Z. Ali, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces enhanced NOMA D2D communications underlaying UAV networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Kuala Lumpur, Malaysia, 2023, pp. 2006–2011.
[25] S. Zargari, A. Hakimi, C. Tellambura, and S. Herath, "User scheduling and trajectory optimization for energy-efficient IRS-UAV networks with SWIPT," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 1815–1830, Feb. 2023.
[26] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042–2055, Jul. 2021.
[27] P. S. Aung, Y. M. Park, Y. K. Tun, Z. Han, and C. S. Hong, "Energy-efficient communication networks via multiple aerial reconfigurable intelligent surfaces: DRL and optimization approach," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 4277–4292, Mar. 2024.
[28] Z. Li, W. Chen, Z. Zhang, Q. Wu, H. Cao, and J. Li, "Robust sum-rate maximization in transmissive RMS transceiver-enabled SWIPT NETWORKS," *IEEE Internet Things J.*, vol. 10, no. 8, pp. 7259–7271, Apr. 2023.
[29] W. Tang et al., "MIMO transmission through reconfigurable intelligent surface: System design, analysis, and implementation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2683–2699, Nov. 2020.
[30] X. Bai et al., "High-efficiency transmissive programmable metasurface for multimode OAM generation," *Adv. Opt. Mater.*, vol.8, no. 17, 2020, Art. no. 2000570.
[31] Z. Li, H. Hu, J. Zhang, and J. Zhang, "Coverage analysis of multiple transmissive RIS-aided outdoor-to-indoor mmWave networks," *IEEE Trans. Broadcast.*, vol. 68, no. 4, pp. 935–942, Dec. 2022.
[32] Z. Li, W. Chen, and H. Cao, "Beamforming design and power allocation for transmissive RMS-based transmitter architectures," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 53–57, Jan. 2022,
[33] Z. Li et al., "Towards transmissive RIS transceiver enabled uplink communication systems: Design and optimization," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 6788–6801, Feb. 2024, doi: 10.1109/JIOT.2023.3312776.
[34] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Performance optimization for UAV-enabled wireless communications under flight time constraints," in *Proc. IEEE Global Commun. Conf.*, Singapore, 2017, pp. 1–6.
[35] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Wireless communication using unmanned aerial vehicles (UAVs): Optimal transport theory for hover time optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8052–8066, Dec. 2017.
[36] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.
[37] M. Asif, A. Ihsan, W. U. Khan, Z. Ali, S. Zhang, and S. X. Wu, "Energy-efficient beamforming and resource optimization for STAR-IRS enabled hybrid-NOMA 6G communications," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 3, pp. 1356–1368, Sep. 2023.
[38] W. Ni, X. Liu, Y. Liu, H. Tian, and Y. Chen, "Resource allocation for multi-cell IRS-aided NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4253–4268, Jul. 2021.
[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ., 2004.
[40] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.
[41] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.
[42] M. Hua, Q. Wu, W. Chen, Z. Fei, H. C. So, and C. Yuen, "Intelligent reflecting surface-assisted localization: Performance analysis and algorithm design," *IEEE Wireless Commun. Lett.*, vol. 13, no. 1, pp. 84–88, Jan. 2024.
[43] C. Ozturk, M. F. Keskin, V. Sciancalepore, H. Wymeersch, and S. Gezici, "RIS-aided localization under pixel failures," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 8314–8329, Aug. 2024.
[44] N. H. Chu, D. T. Hoang, D. N. Nguyen, N. Van Huynh, and E. Dutkiewicz, "Joint speed control and energy replenishment optimization for UAV-assisted IoT data collection with deep reinforcement transfer learning," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 5778–5793, Apr. 2023.
[45] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Optimal transport theory for cell association in UAV-enabled cellular networks," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2053–2056, Sep. 2017.
[46] A. A. Hammadi, L. Bariah, S. Muhaidat, M. Al-Qutayri, P. C. Sofotasios, and M. Debbah, "Deep Q-learning-based resource allocation in NOMA visible light communications," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 2284–2297, 2022.
[47] A. Benfaid, N. Adem, and B. Khalfi, "AdaptSky: A DRL based resource allocation framework in NOMA-UAV networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, 2021, pp. 1–7.