




Please cite the Published Version

Umer, Muhammad , Mohsin, Muhammad Ahmed , Kaushik, Aryan , Nadeem, Qurrat-UI-Ain , Nasir, Ali Arshad  and Hassan, Syed Ali  (2025) Reconfigurable Intelligent Surface-Assisted Aerial Nonterrestrial Networks: An Intelligent Synergy With Deep Reinforcement Learning. IEEE Vehicular Technology Magazine. ISSN 1556-6080

DOI: <https://doi.org/10.1109/mvt.2024.3524745>

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/638224/>







Usage rights:  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

Additional Information: This is an author-produced version of the published paper. Uploaded in accordance with the University's Research Publications Policy.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

RIS-Assisted Aerial Non-Terrestrial Networks: An Intelligent Synergy with Deep Reinforcement Learning

Muhammad Umer , Muhammad Ahmed Mohsin , Aryan Kaushik , Qurrat-Ul-Ain Nadeem ,
Ali Arshad Nasir , and Syed Ali Hassan 

Abstract—Reconfigurable intelligent surface (RIS)-assisted aerial non-terrestrial networks (NTNs) offer a promising paradigm for enhancing wireless communications in the era of 6G and beyond. By integrating RIS with aerial platforms such as unmanned aerial vehicles (UAVs) and high-altitude platforms (HAPs), these networks can intelligently control signal propagation, extending coverage, improving capacity, and enhancing link reliability. This article explores the application of deep reinforcement learning (DRL) as a powerful tool for optimizing RIS-assisted aerial NTNs. We focus on hybrid proximal policy optimization (H-PPO), a robust DRL algorithm well-suited for handling the complex, hybrid action spaces inherent in these networks. Through a case study of an aerial RIS (ARIS)-aided coordinated multi-point non-orthogonal multiple access (CoMP-NOMA) network, we demonstrate how H-PPO can effectively optimize the system and maximize the sum rate while adhering to system constraints. Finally, we discuss key challenges and promising research directions for DRL-powered RIS-assisted aerial NTNs, highlighting their potential to transform next-generation wireless networks.

I. INTRODUCTION

Sixth-generation (6G) wireless networks promise ubiquitous and seamless connectivity, catering to the ever-growing demands of an increasingly interconnected world. With the rapid growth of data-intensive and delay-sensitive applications, such as extended reality, autonomous driving, and the Internet of Things (IoT), existing terrestrial networks face significant challenges in terms of capacity, coverage, latency, and efficiency [1]. This necessitates a paradigm shift in network design towards non-terrestrial networks (NTNs), specifically aerial NTNs, which leverage a constellation of aerial platforms, including satellites and high-altitude platforms (HAPs), to augment and extend terrestrial network capabilities.

As envisioned by the Third Generation Partnership Project (3GPP) and the International Mobile Communication (IMT)-2030 framework, aerial NTNs will play a pivotal role in achieving the ambitious connectivity goals of 6G and beyond by providing resilient and sustainable communication infrastructure. Unmanned aerial vehicles (UAVs) are a key component of aerial NTNs, offering enhanced positioning

freedom, cost-effective deployment and maintenance, and the ability to establish strong line-of-sight (LoS) links. UAVs can operate as aerial base stations (ABSs), aerial relays (ARs), or aerial user equipment (AUEs), each contributing to enhanced network performance through a variety of use cases [2], [3].

Complementing the flexibility of NTNs is the transformative technology of reconfigurable intelligent surfaces (RISs). RISs are engineered surfaces comprising a large number of passive reflecting elements that can intelligently manipulate the propagation of electromagnetic waves. By dynamically controlling the phase shifts of these elements, RISs can enhance desired signals, suppress interference, and reshape the wireless channel to improve communication quality, offering promising applications in coverage extension, interference mitigation, and physical layer security enhancement [4], [5]. However, traditional, fixed terrestrial RIS (TRIS) deployments often face limitations in placement and reflection angles. Mounting RIS on aerial platforms to form aerial RIS (ARIS) overcomes these limitations. ARIS leverages the mobility of aerial vehicles to achieve dynamic positioning and their altitude to achieve panoramic full-angle reflection capabilities, thereby optimizing signal reflection and maximizing communication performance [6]. This flexibility allows ARIS to adapt effectively to changing channel conditions, user distributions, and environmental factors.

Optimizing the performance of ARIS requires sophisticated control mechanisms to effectively manage the complex interplay of RIS configurations, available resources, trajectories of aerial platforms, and dynamic channel conditions [7]. Deep reinforcement learning (DRL) emerges as a powerful tool to address these complexities. DRL algorithms can learn optimal control policies through trial and error, adapting to changing environments and maximizing long-term performance objectives. Using deep neural networks, DRL can handle high-dimensional state and action spaces, making it particularly well-suited for the intricate optimization problems inherent in NTNs [8]. Therefore, this article explores the application and ability of DRL to enhance network performance in RIS-aided aerial NTNs.

The rest of the article is organized as follows. The following section provides an overview of RIS-assisted communications in the context of aerial NTNs. We then motivate the usage of DRL for network optimization and present a detailed description of proximal policy optimization (PPO). Next, a case study showcasing the effectiveness of hybrid PPO (H-PPO) in an ARIS-aided coordinated multi-point non-orthogonal multiple

Muhammad Umer, Muhammad Ahmed Mohsin, and Syed Ali Hassan are with the School of Electrical Engineering and Computer Science (SEECs), National University of Sciences and Technology (NUST), Pakistan.

Aryan Kaushik is with the Department of Computing and Mathematics, Manchester Metropolitan University, UK.

Qurrat-Ul-Ain Nadeem is with New York University (NYU) Abu Dhabi, UAE, and NYU Tandon School of Engineering, USA.

Ali Arshad Nasir is with the King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia.

access (CoMP-NOMA) system is presented. We conclude by discussing the key challenges and emerging research directions for DRL-powered ARIS in the rapidly evolving landscape of future wireless networks.

II. RIS-AIDED AERIAL NTN: AN OVERVIEW

We commence by providing an overview of RIS-assisted communications within the context of aerial NTNs, highlighting their potential to revolutionize wireless networks by exploring their advantages, key optimization aspects, and potential use cases.

A. Advantages of Aerial RIS

ARIS presents significant advantages over TRIS due to its deployment capabilities and better performance characteristics. Unlike TRIS, which is confined to fixed locations, ARIS can be integrated with various aerial platforms, including UAVs, HAPs, and even satellites, enabling flexible and dynamic deployment. This mobility allows for on-demand coverage extension, rapid deployment in disaster scenarios, and adaptive positioning for optimal signal reflection. Moreover, the aerial nature of ARIS facilitates 360° panoramic full-angle reflection, surpassing the typical 180° half-space limitation of TRIS mounted on a building facade or any other vertical surface.

The elevated positioning of ARIS also leads to several substantial improvements in communication performance. Strategic aerial positioning enables ARIS to reflect signals in a way that establishes LoS links, leading to a much lower probability of signal blockage. This implies better channel quality, resulting in improved data rates and increased reliability in communication links between satellites, aerial platforms, and ground users [9]. The combination of deployment flexibility and superior channel characteristics makes ARIS a compelling technology for future wireless networks, particularly in scenarios requiring dynamic coverage optimization or rapid network deployment.

B. An Optimization Perspective

Optimizing the performance of RIS-assisted aerial NTNs requires a coordinated approach that considers the unique characteristics of both the RIS and aerial platforms.

1) *Passive Beamforming*: RIS utilizes passive beamforming to enhance desired signals and mitigate interference. This involves dynamically adjusting the phase shifts of the reflecting elements on the RIS to constructively combine desired signals at the receiver while suppressing undesired signals. Thus, the objective is to maximize the signal-to-interference-plus-noise ratio (SINR) and achieve higher data rates.

2) *Trajectory Control*: The trajectories of aerial platforms equipped with RIS need careful optimization to maximize coverage, minimize path loss, and avoid obstacles. Trajectory control involves determining the optimal flight paths, altitudes, and orientations of UAVs or HAPs to ensure efficient signal reflection and coverage for users. Factors such as user distribution, channel conditions, energy efficiency, and airspace regulations must be considered [3].

3) *Resource Allocation*: Realizing the performance gains of ARIS-assisted aerial NTNs necessitates efficient resource allocation, which entails strategically assigning communication resources, such as power, bandwidth, and time slots, to different network entities [7]. Factors like user demand, channel conditions, quality of service (QoS) requirements, and energy constraints need to be considered for efficient resource utilization.

C. Use Cases of RIS in Aerial NTNs

The unique capabilities of RIS integrated with aerial NTNs make it suitable for various use cases, as illustrated in Fig. 1.

1) *Ubiquitous Connectivity for Remote IoT*: ARIS, by leveraging the flexibility and mobility of aerial platforms like UAVs and HAPs, can extend connectivity to remote IoT devices, even in areas lacking terrestrial infrastructure [10]. As shown in Fig. 1, a BS transmits a signal to a LEO satellite, which relays it via inter-satellite links to another RIS-equipped LEO positioned over the remote area. This LEO reflects the signal to an RIS mounted on a HAP, which directs it to the IoT devices on the ground. This ARIS and satellite communication-enabled multi-hop link provides reliable, low-latency connectivity to remote areas, supporting crucial applications like smart agriculture and environmental monitoring.

2) *Enhanced Urban Coverage*: Dense urban environments often suffer from signal blockage and multipath fading, hindering communication reliability and data rates. ARIS can be strategically deployed to overcome these challenges, as shown in Fig. 1. By reflecting signals from the BS or LEO satellites, ARIS creates alternative signal paths, bypassing obstacles like tall buildings and extending coverage to shadowed areas. This dynamic positioning enables improved signal quality and extended connectivity for users in challenging urban environments.

3) *High-Capacity Hotspots*: High-density user scenarios, such as stadiums or concert venues, require high-capacity connectivity to meet the simultaneous data needs of numerous users. ARIS can effectively address these demands by intelligently reflecting and directing signals from BSs or other aerial platforms towards the high-density area. Fig. 1 depicts multiple ARISs hovering over a stadium, reflecting and focusing signals from a BS to provide high-quality, high-capacity connectivity to the dense user population.

4) *Secure UAV Swarms*: Secure and reliable communication is vital for UAV swarm operations, especially in sensitive applications like surveillance, data collection, and disaster response [11]. ARIS can significantly enhance the security of UAV swarm communications by mitigating jamming and eavesdropping attempts. Strategically positioned ARIS can create focused beams toward the intended UAV swarm while simultaneously creating nulls in the direction of potential eavesdroppers or jammers. Fig. 1 shows how two strategically positioned ARISs can protect a UAV swarm. One ARIS amplifies the desired signal from the BS towards the swarm, ensuring reliable communication, while the second ARIS reflects a jamming signal from a malicious drone towards an

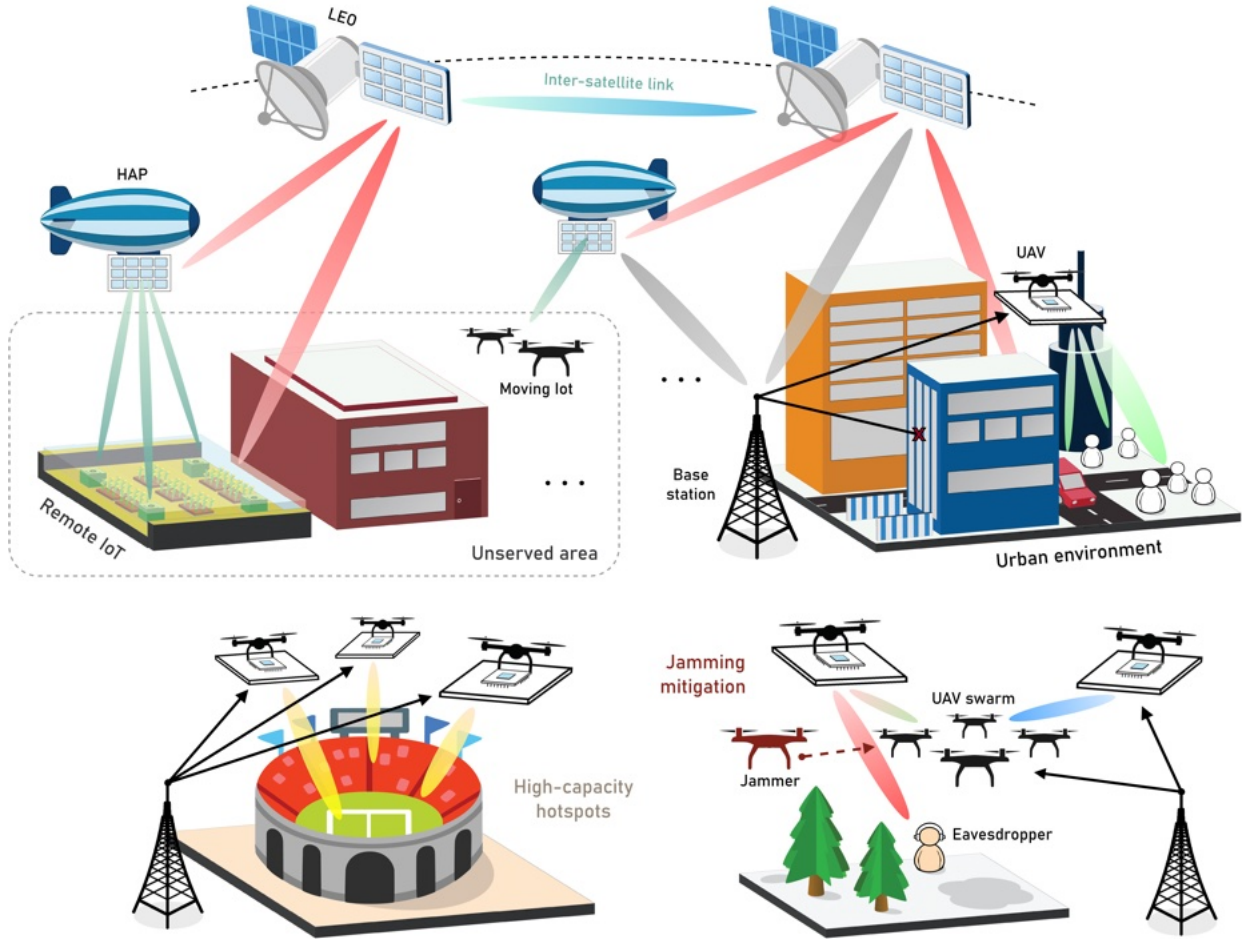


Fig. 1. Illustrative use cases of RIS in aerial NTN for enhanced connectivity, coverage, and security.

eavesdropper, thereby protecting the swarm’s communication integrity.

III. DEEP REINFORCEMENT LEARNING FOR ENHANCED RIS-ASSISTED AERIAL NTN

In this section, we explore the application of DRL for optimizing RIS-assisted aerial NTN communication. We motivate the use of DRL, provide a detailed explanation of a state-of-the-art DRL algorithm, and present a case study showcasing the effectiveness of the proposed DRL-based solution.

A. Why DRL?

Reinforcement learning (RL), in essence, is the science of decision making. In contrast to supervised learning, which relies on labeled data, RL involves an agent learning to make decisions through trial and error, interacting with an environment and receiving rewards or penalties based on its actions. The ultimate goal of any RL agent is to learn a policy that maximizes its cumulative reward over time. This makes RL particularly suitable for dynamic and complex systems, where it is difficult or infeasible to pre-program optimal behavior.

Deep reinforcement learning (DRL) enhances RL by incorporating deep neural networks as function approximators

which are capable of handling high-dimensional state and action spaces, such as those found in complex wireless communication systems. To understand the learning behaviour of RL agents, it is essential to introduce the concepts of state value, state-action value, and policy.

- *State Value*: The value of a state represents the expected long-term reward the agent can achieve starting from that state and following a specific policy. It quantifies the *goodness* of being in a particular state.
- *State-Action Value*: The state-action value represents the expected long-term reward when starting in a specific state, taking a particular action, and then following the given policy.
- *Policy*: A policy, denoted by π_d for discrete policies and π_c for continuous policies, is a function that maps states to actions. It dictates the behaviour of agent, guiding it to choose actions based on the observed state aiming to maximize the expected long-term reward.

Optimizing RIS-assisted aerial NTN presents unique challenges due to the inherent complexity and dynamic nature of these systems. The interplay of satellite movement, aerial platform trajectories, RIS configurations, and resource allocation strategies across high-dimensional state and action spaces demands intelligent and adaptive control mechanisms.

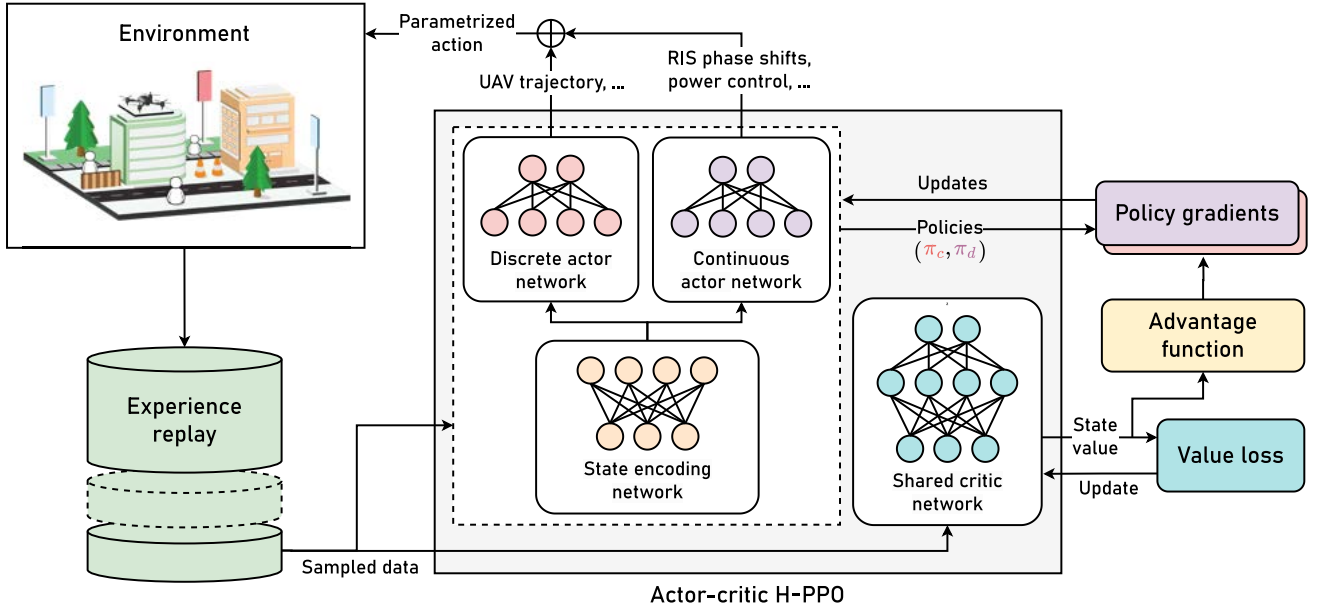


Fig. 2. Architecture of H-PPO for optimization of RIS-assisted aerial NTNs.

While conventional optimization techniques, such as convex optimization, have been applied to wireless communication problems, they often struggle to cope with the dynamic and unpredictable nature of RIS-assisted aerial NTNs. These techniques typically rely on accurate and instantaneous channel state information (CSI) and involve solving computationally intensive optimization problems, leading to significant overhead and delays. This is particularly problematic in scenarios with mobile UAVs, where frequent updates to CSI and resource allocation are necessary, further amplifying the computational burden and impacting real-time performance.

DRL emerges as a powerful solution for addressing these challenges. Unlike traditional optimization approaches, DRL algorithms can efficiently learn and update control policies online, adapting to the dynamic nature of these networks where real-time adaptation is crucial. Among the various DRL algorithms, PPO stands out as a highly effective choice. As a classical policy gradient algorithm, PPO exhibits enhanced stability and adaptability, making it well-suited for navigating the challenges posed by the dynamic environments of RIS-assisted aerial NTNs. This adaptability stems from its ability to adjust the policy update step size during training, contrasting with conventional policy gradient algorithms that rely on a fixed and often challenging-to-tune step size. PPO further enhances stability and efficiency through several key mechanisms, which will be explored in detail in the following subsection. Additionally, we introduce H-PPO, an extension of PPO specifically tailored for hybrid action spaces, a common characteristic of RIS-assisted aerial NTNs.

B. Description of PPO & H-PPO

PPO is a policy gradient-based DRL algorithm acclaimed for its simplicity, reliability, and efficiency. Unlike traditional policy gradient methods, which often suffer from instability due to large updates, PPO aims to improve the policy in a

more controlled manner; policy updates in PPO are bounded by a *trust region* enacted by a surrogate objective function, preventing drastic changes that could destabilize the learning process.

PPO achieves stable and efficient learning through several key mechanisms. To mitigate the high variance associated with traditional policy gradient methods, PPO employs a *clipped surrogate objective function*, ensuring stability by keeping updates within a defined trust region achieved through clipping of the probability ratio between new and old policies. PPO further boosts data efficiency by performing *multi-epoch updates* on each sampled batch, allowing the agent to extract more knowledge from experiences and accelerate overall learning. Furthermore, PPO makes use of the *advantage function*, which estimates the relative benefit of taking a specific action compared to the average action at a given state. By prioritizing actions with higher advantages, PPO focuses on learning from the most rewarding experiences, leading to faster convergence towards an optimal policy. Lastly, PPO employs *experience replay*, a common technique in DRL, where past interactions are stored in a memory buffer and randomly sampled during training. This reduces data correlation and improves learning stability by preventing the algorithm from becoming overly biased towards recent experiences.

Traditional RL algorithms, such as deep Q-networks (DQN) and deep deterministic policy gradient (DDPG), are primarily designed for either purely discrete or purely continuous action spaces. However, in the context of RIS-assisted aerial NTNs, we need to control both discrete actions, such as aerial maneuvers, and continuous actions, like precise adjustments of power control levels and RIS phase shifts.

Applying DQN directly to this scenario would require discretizing the continuous actions, resulting in an impractically large action space that would hinder convergence and make learning inefficient. Similarly, while DDPG can handle

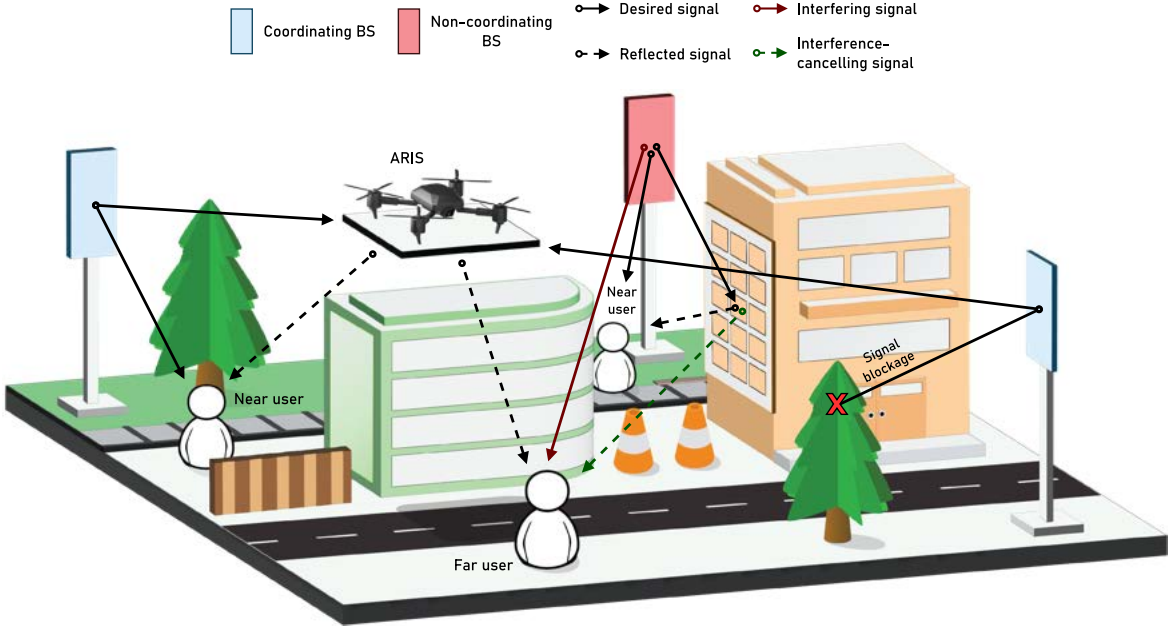


Fig. 3. System model of ARIS-assisted CoMP-NOMA network.

continuous actions, it may not be ideal for the hybrid action spaces in RIS-assisted aerial NTN. DDPG can also exhibit instability when dealing with the complex, non-linear relationships between trajectories, RIS configurations, and communication performance, especially in dynamic environments with changing channel conditions.

To overcome this challenge, we employ H-PPO, as illustrated in Fig. 2. H-PPO extends the standard PPO framework by incorporating multiple output heads, allowing for simultaneous optimization of both discrete and continuous actions without resorting to excessive discretization. A shared critic network provides a common performance benchmark for both discrete and continuous actions by evaluating the value function for all states. The state encoding network processes the input state information from the sampled batch, creating a shared representation that is then fed to two separate actor heads: one dedicated to discrete actions and the other to continuous actions. While both actors interact with the same environment, their optimization occurs independently. Each actor utilizes its own distinct objective function, resulting in separate policy gradients tailored to its specific action type.

C. Case Study: DRL-Enabled ARIS-Assisted CoMP-NOMA

Building upon the evolution of cellular networks, coordinated multi-point (CoMP) techniques have been standardized to address inter-cell interference and spectrum limitations. Non-orthogonal multiple access (NOMA) further enhances spectral efficiency by allowing multiple users to share the same time-frequency resources through superposition coding of user signals at the transmitter and successive interference cancellation (SIC) at the receivers. CoMP-NOMA networks offer synergistic advantages by combining interference management through CoMP with spectral efficiency gains from NOMA. When paired with ARIS, it presents a compelling scenario for

next-generation wireless networks with significant potential for enhancing coverage, capacity, and reliability.

However, such an integrated system presents a complex optimization problem due to the inherent coupling of control variables, including the trajectory of aerial platforms, RIS phase shifts, power allocation factors, and dynamic channel conditions. Conventional methods often struggle to effectively handle such coupled parameters, particularly in the context of real-time adaptation to changing wireless environments. As discussed earlier, DRL-based optimization provides an efficient and adaptable approach to managing such scenarios. The ability to learn and adapt to dynamic systems, handle high-dimensional state and action spaces, and optimize without explicit mathematical models makes it an excellent choice for addressing these challenges.

In this subsection, we demonstrate how H-PPO effectively manages the complexities of an ARIS-aided CoMP-NOMA network and how it can optimally allocate resources to maximize network sum rate.

1) *System Description:* As illustrated in Fig. 3, we consider a downlink CoMP-NOMA network with both TRIS and ARIS. Three BSs are present, two of which utilize CoMP to serve a far user (FU) and are aided by the ARIS, whereas the third non-CoMP BS transmits to its own near user (NU), generating interference for the FU. We assume that direct links from the coordinating BSs to the FU are obstructed by obstacles, emphasizing the critical role of ARIS in establishing reliable communication. The signal from the non-CoMP BS to the ARIS is neglected due to the double path loss inherent in reflection links and the substantial propagation distance between them. This setup emphasizes the practical challenge of serving users in obstructed or shadowed locations while also demonstrating ARIS' ability to establish and maintain dynamic LoS links.

All communication channels in the network are modeled

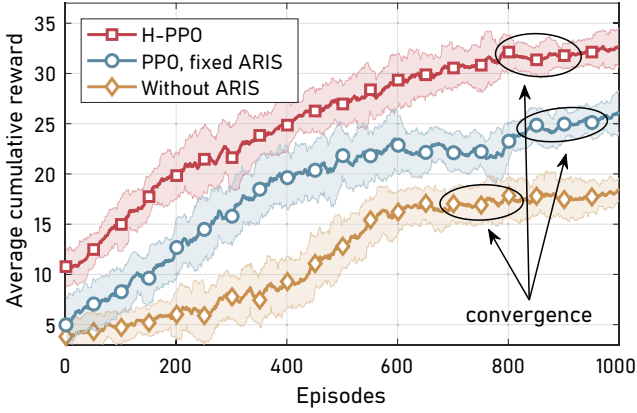


Fig. 4. Average cumulative reward versus the number of training episodes.

using the Nakagami- m fading distribution with varying fading parameter m , providing a flexible representation of diverse channel conditions. Without loss of generality, we assume that the NUs experience better channel conditions than the FU, justifying the application of NOMA. For simplicity, we assume perfect CSI knowledge is available at the central controller, which acts as the DRL agent in this scenario. However, it should be noted that imperfect CSI due to factors such as ARIS jittering, atmospheric effects, and estimation errors is a crucial challenge for practical deployments.

The objective is to maximize the cumulative reward, defined as the network sum rate with penalties for violating operational constraints, over the UAV's operational time, which is discretized into time slots. To achieve this, the agent learns to control the UAV's trajectory, the phase shifts of both the ARIS and TRIS, and the NOMA power allocation factors. This coordinated control is subject to constraints that ensure the proper operation of SIC, maintain the UAV within a designated area of interest, and keep the phase shifts of both RISs within practical bounds.

2) *Performance Evaluation*: To evaluate the effectiveness of H-PPO for optimizing the ARIS-assisted CoMP-NOMA network, we conduct simulations using the following parameters unless stated otherwise. Each of the three BSs has a transmit power of 15 dBm and serves user equipment distributed on a grid of 150×150 m². The ARIS is initially positioned at the center of this grid. Additionally, the system is assumed to be operating with a bandwidth of 10 MHz at a carrier frequency of 2.4 GHz.

In Fig. 4, we evaluate the average cumulative reward achieved by the DRL agent across training episodes, showcasing the convergence behavior of various DRL configurations. Both PPO and H-PPO, under different system configurations, demonstrate stable convergence, reaching a plateau in cumulative reward as training progresses. Notably, H-PPO with coordinated phase shift control of both ARIS and TRIS achieves the highest average cumulative reward. This underscores the effectiveness of jointly optimizing UAV trajectory, and passive beamforming to maximize network sum rate. Moreover, the H-PPO configuration outperforms PPO with a fixed ARIS and PPO without ARIS, highlighting the importance of dynamic

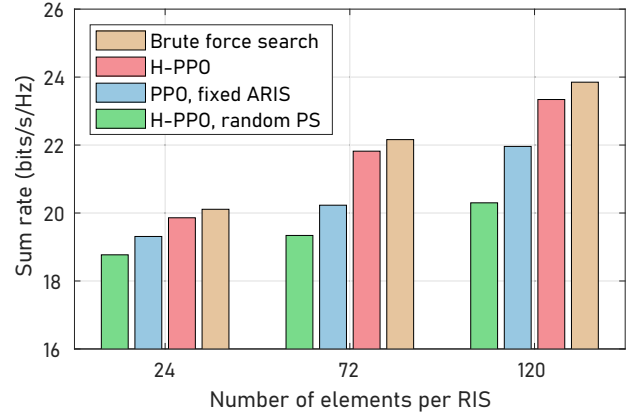


Fig. 5. Network sum rate reward versus the number of elements in both ARIS and TRIS.

ARIS positioning for establishing optimal LoS links.

Fig. 5 illustrates the impact of the number of reflecting elements in both ARIS and TRIS on the achievable network sum rate. To benchmark the DRL algorithms against the optimal solution, we perform a brute-force search over all possible combinations of UAV positions and RIS phase shifts, which provides a performance upper bound. As the number of elements increases, the achievable sum rate generally improves due to enhanced beamforming capabilities. As can be observed, H-PPO achieves near-optimal performance but deviates from the optimal solution as the number of elements increases due to the larger action space, which is harder for DRL agents to optimize. Comparatively, PPO with fixed ARIS and H-PPO with random ARIS phase shifts lag behind H-PPO.

Fig. 6 shows the impact of reward function design on the network sum rate. We explored the following reward function designs.

- *Sum rate reward*: Though intuitive for maximizing sum rate, this approach leads to suboptimal performance. The agent prioritizes momentarily high sum rates but may violate operational constraints, like the UAV leaving the designated area or assigning power allocation factors that hinder SIC.
- *Penalized sum rate reward*: This design incorporates penalties for constraint violations and encourages the agent to balance sum rate maximization with constraint adherence.
- *Multi-objective rewards*: Agent optimizes multiple objectives simultaneously, each with a distinct reward function. This allows exploration of trade-offs between conflicting goals, such as achieving high sum rates while maintaining user fairness.
- *Compound rewards*: Unlike multi-objective rewards, this design combines multiple reward signals into a single function through weighted sums, which simplifies the learning process.

Interestingly, the highest sum rate is achieved by a compound reward function that integrates sum rate, energy efficiency, and UAV trajectory stability into a single reward signal, suggesting that holistic reward designs can be more effective.

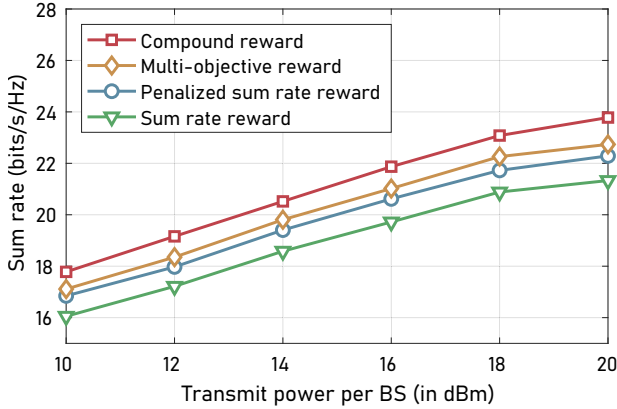


Fig. 6. Network sum rate versus transmit power per BS for different reward functions.

D. Complexity Analysis

Computational complexity is of critical importance for any algorithm, as it dictates its eventual real-world deployment. For H-PPO and DRL algorithms in general, the complexity is largely determined by the size of the neural networks. The overall complexity can be approximated as $\mathcal{O}[n^2(Q_s + Q_d + Q_c)]$, where n represents the number of neurons in each layer of the state encoding, discrete, and continuous actor networks, and Q_s , Q_d , and Q_c denote the number of layers in each respective network. The complexity increases quadratically with the number of neurons and linearly with the number of layers. The actual computational cost depends on various factors, including hardware specifications, software implementations, and hyperparameter settings.

Mathematically analyzing the convergence of H-PPO is challenging due to the inherent complexities of neural networks and their dependence on various parameters. However, we empirically verified convergence by monitoring the average cumulative reward over numerous training episodes, as shown in Fig. 4. Stable convergence, indicating successful learning, was observed as the reward reached a plateau. This validation provides confidence in the effectiveness of H-PPO in learning near-optimal policies in a complex environment, such as the one in our case study. It is important to note that the convergence speed drastically depends on the choice of hyperparameters, such as the learning rate, clipping parameter, and discount factor, and may require extensive tuning to achieve optimal performance.

IV. CHALLENGES AND FUTURE DIRECTIONS

While RIS-assisted aerial NTN hold immense potential, numerous challenges and future research directions need to be addressed to fully unlock their capabilities and enable widespread adoption in next-generation wireless networks.

A. Algorithmic Challenges

The application of DRL to optimize RIS-assisted aerial NTN presents several algorithmic challenges that require further research. Model-free DRL algorithms like DDPG and PPO often exhibit sample inefficiency, particularly in complex

scenarios like aerial NTN where they require numerous interactions with the environment to learn effective state-to-action mappings. Improving sample efficiency through meta-learning, transfer learning, or model-based RL offers a promising direction for enabling faster training and adaptability [12].

Designing effective reward functions remains another major challenge. A reward function must balance multiple, often conflicting objectives such as sum rate maximization, power minimization, user fairness, and constraint satisfaction. This typically requires many iterations of trial and error. Techniques like intrinsic motivation or curiosity-driven exploration could improve exploration strategies, especially for high-dimensional action spaces. Additionally, DRL policies trained for specific scenarios often struggle to generalize to new environments with different user distributions, channel conditions, platform configurations, or even minor changes in system parameters. Consequently, scalability to larger networks with multiple ARIS and diverse channel conditions becomes increasingly difficult. Developing agents with improved generalization capabilities and effective transfer learning mechanisms is crucial for ensuring adaptability and scalability of DRL-powered solutions in real-world deployments.

B. Implementation Challenges

Several system-level challenges hinder the implementation of ARIS in real-world scenarios. A fundamental issue lies in the accurate control of the ARIS platform and its elements, particularly due to the inherent instability of non-terrestrial networks and environmental perturbations [13]. Constant micro-movements and vibrations lead to phase shift errors and misalignment of RIS elements. When combined with the highly dynamic nature of aerial platforms, this introduces Doppler effects and time-varying channel conditions that complicate control mechanisms.

Imperfect CSI from estimation errors, quantization effects, and feedback delays can severely degrade overall network performance. Jittering motion of aerial platforms and atmospheric conditions further exacerbate these imperfections, introducing additional uncertainties in channel estimation. Real-world deployments must also be resilient to adversarial attacks, necessitating the development of robust beamforming algorithms and mechanisms that address security vulnerabilities in RIS control signaling and communication links. These systems must be capable of mitigating susceptibility to jamming and eavesdropping, and strong encryption techniques and sophisticated intrusion detection systems could provide the necessary safeguards.

Power consumption and energy efficiency remain major limiting factors as well. Although RIS elements themselves are passive, the overall energy expenditure of ARIS-assisted aerial NTN is significant, especially for battery-powered UAVs. Careful planning is required for power consumption related to propulsion, control, onboard signal processing, and data transmission. Developing algorithms for optimal power allocation strategies and energy-efficient communication is essential for maximizing operational time and extending the service region of UAVs in real-world applications.

C. Emerging Directions

To advance ARIS capabilities and expand their applications, several promising research directions seem particularly compelling. Integrating RIS-assisted aerial NTN with edge computing would bring computation and storage resources closer to users, facilitating low-latency processing, localized data management, and real-time data analytics for applications like autonomous driving, drone traffic control, and distributed AI [14]. This integration would support the increasing demands of data-intensive and latency-sensitive applications. Distributed learning techniques like federated learning could enable collaborative learning across multiple RIS-equipped platforms and users while protecting data privacy and enhancing scalability in large-scale deployments. This approach offers a promising way to manage the increased complexity and diverse environmental conditions encountered in vast NTN deployments by enabling collaborative learning and adaptation among distributed agents.

ARIS in integrated sensing and communication (ISAC) systems could synergistically optimize both communication and sensing functionalities and improve performance to address inherent trade-offs [15]. Lastly, the development of active ARIS (AARIS), which incorporates amplifiers into reflecting elements, can significantly enhance coverage, capacity, and reliability. Although active metasurfaces offer promising potential to overcome limitations of double path loss, the active components would introduce new challenges in energy and power management.

V. CONCLUSION

This article presented an overview of RIS-assisted aerial NTNs, highlighting their transformative potential in shaping the future of wireless communications. By combining the flexibility of aerial platforms with the intelligent signal manipulation capabilities of RIS, these networks offer compelling solutions to challenges in 6G and beyond. We motivated the use of DRL as a powerful tool for optimizing the complex interactions in these dynamic networks. Moreover, we presented PPO and its multi-output extension, H-PPO, as effective DRL algorithms capable of handling the hybrid action spaces often encountered in RIS-assisted aerial NTNs. A case study on an ARIS-aided CoMP-NOMA network demonstrated the superior performance of H-PPO in maximizing network sum rate.

While challenges persist in practical implementation in large-scale networks, channel estimation, robustness, and security, ongoing research actively explores solutions for implementing DRL-powered RIS-assisted aerial NTNs. Scaling DRL algorithms to manage multiple RIS-equipped aerial platforms across diverse environmental conditions necessitates further research, particularly in distributed learning and control mechanisms.

REFERENCES

- [1] L.-H. Shen, K.-T. Feng, and L. Hanzo, "Five facets of 6G: Research challenges and opportunities," *ACM Comput. Surv.*, vol. 55, no. 11, 2023.
- [2] Y. Li, Y. Bi, J. Wang, Z. Li, H. Zhang, and P. Zhang, "Unmanned aerial vehicle assisted communication: Applications, challenges, and future outlook," *Clust. Comput.*, vol. 27, no. 9, pp. 13187–13202, 2024.

- [3] Y. Bai, H. Zhao, X. Zhang, Z. Chang, R. Jäntti, and K. Yang, "Toward autonomous multi-UAV wireless network: A survey of reinforcement learning-based approaches," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 4, pp. 3038–3067, 2023.
- [4] E. Basar, G. C. Alexandropoulos, Y. Liu, Q. Wu, S. Jin, C. Yuen, O. A. Dobre, and R. Schober, "Reconfigurable intelligent surfaces for 6G: Emerging hardware architectures, applications, and open challenges," *IEEE Veh. Technol. Mag.*, vol. 19, no. 3, pp. 27–47, 2024.
- [5] F. Wang and A. L. Swindlehurst, "Applications of absorptive reconfigurable intelligent surfaces in interference mitigation and physical layer security," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 5, pp. 3918–3931, 2024.
- [6] X. Gu, W. Duan, G. Zhang, M. Wen, J. Choi, and P.-H. Ho, "Aerial reconfigurable intelligent surface-assisted terrestrial communications," *IEEE Internet Things Mag.*, vol. 7, no. 2, pp. 54–60, 2024.
- [7] A. C. Pogaku, D.-T. Do, B. M. Lee, and N. D. Nguyen, "UAV-assisted RIS for future wireless communications: A survey on optimization and performance analysis," *IEEE Access*, vol. 10, pp. 16320–16336, 2022.
- [8] S. Mahboob and L. Liu, "Revolutionizing future connectivity: A contemporary survey on AI-empowered satellite-based non-terrestrial networks in 6G," *IEEE Commun. Surv. Tutor.*, vol. 26, no. 2, pp. 1279–1321, 2024.
- [9] M. Toka, B. Lee, J. Seong, A. Kaushik, J. Lee, J. Lee, N. Lee, W. Shin, and H. V. Poor, "RIS-empowered LEO satellite networks for 6G: Promising usage scenarios and future directions," *IEEE Commun. Mag.*, vol. 62, no. 11, pp. 128–135, 2024.
- [10] M. A. Jamshed, A. Kaushik, M. Toka, W. Shin, M. Z. Shakir, S. P. Dash, and D. Dardari, "Synergizing airborne non-terrestrial networks and reconfigurable intelligent surfaces-aided 6G IoT," *IEEE Internet Things Mag.*, vol. 7, no. 2, pp. 46–52, 2024.
- [11] S. Javed, A. Hassan, R. Ahmad, W. Ahmed, R. Ahmed, A. Saadat, and M. Guizani, "State-of-the-art and future research challenges in UAV swarms," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19023–19045, 2024.
- [12] S. Faramarzi, S. Javadi, F. Zeinali, H. Zarini, M. R. Mili, M. Bennis, Y. Li, and K.-K. Wong, "Meta reinforcement learning for resource allocation in aerial active-RIS-assisted networks with rate-splitting multiple access," *IEEE Internet Things J.*, pp. 1–1, 2024.
- [13] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, "Channel estimation with reconfigurable intelligent surfaces—a general framework," *Proc. IEEE*, vol. 110, no. 9, pp. 1312–1338, 2022.
- [14] B. Shang, H. V. Poor, and L. Liu, "Aerial reconfigurable intelligent surfaces meet mobile edge computing," *IEEE Wirel. Commun.*, vol. 29, no. 6, pp. 104–111, 2022.
- [15] Z. Wu, X. Li, Y. Cai, and W. Yuan, "Joint trajectory and resource allocation design for RIS-assisted UAV-enabled ISAC systems," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 5, pp. 1384–1388, 2024.

Muhammad Umer (mumer.bec20seecs@seecs.edu.pk) received the B.E. degree in electrical engineering from National University of Sciences and Technology (NUST), Pakistan. His current research interests include coordinated multi-point (CoMP) transmission, non-orthogonal multiple access (NOMA), reconfigurable intelligent surface (RIS), and deep reinforcement learning (DRL).

Muhammad Ahmed Mohsin (muahmed@stanford.edu) received the B.E. degree in electrical engineering from National University of Sciences and Technology (NUST), Pakistan. He is currently pursuing the Ph.D. degree in electrical engineering from Stanford University, USA. His primary focus of research lies in next-generation wireless communications, deep learning (DL), and reinforcement learning (RL).

Aryan Kaushik (a.kaushik@mmu.ac.uk) received the Ph.D. degree in communications engineering from the Institute for Digital Communications, School of Engineering, The University of Edinburgh, UK, in 2020. He is currently an Associate Professor with the Department of Computing and Mathematics, Manchester Metropolitan University, UK. His research interests include signal processing for 5G/6G wireless communications, integrated sensing and communications, reconfigurable holographic and intelligent surfaces, energy efficient communication systems, non-terrestrial networks, mmWave massive MIMO, and machine learning, edge computing and AI-assisted communications.

Qurrat-Ul-Ain Nadeem (qurrat.nadeem@nyu.edu) received the M.S. and Ph.D. degrees in electrical engineering from the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, in 2015 and 2018, respectively. She is currently an Assistant Professor at New York University Abu Dhabi, and NYU WIRELESS, NYU Tandon School of Engineering. Her research interests lie in the areas of wireless communications, signal processing, and electromagnetics and antenna theory.

Ali Arshad Nasir (anasir@kfupm.edu.sa) received his the.D. degree in telecommunications engineering from the Australian National University (ANU), Australia, in 2013. He joined the Department of Electrical Engineering at King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 2016, where he is currently working as an Associate Professor. His research interests are in the area of signal processing in wireless communication systems. He is an Associate Editor for IEEE Communications Letters.

Syed Ali Hassan (ali.hassan@seecs.edu.pk) received the Ph.D. degree in electrical engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, USA, in 2011, an M.S. in mathematics from Georgia Tech in 2011, and an M.S. in electrical engineering from the University of Stuttgart, Germany, in 2007. Currently, he is a Professor with the School of Electrical Engineering and Computer Science (SEECS), NUST, where he is the Director of the Information Processing and Transmission Lab, which focuses on various aspects of theoretical communications. He has authored more than 300 conference and journal articles, mainly in the area of wireless communications.