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. Furthermore, 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, 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 the 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). A case study

Muhammad Umer, Muhammad Ahmed Mohsin, and Syed Ali Hassan are with the National University of Sciences and Technology (NUST), Pakistan.

Aryan Kaushik is with University of Sussex, 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 King Fahd University of Petroleum and Minerals (KFUPM), Saudia Arabia.

showcasing the effectiveness of hybrid PPO (H-PPO) in an ARIS-aided coordinated multi-point non-orthogonal multiple access (CoMP-NOMA) system is then 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 NTNs: 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 offers several key advantages over TRIS, making them a compelling technology for future wireless networks.

1) Deployment Flexibility: Unlike TRIS, which is typically constrained to fixed locations, RIS can be integrated with various aerial platforms, including UAVs, HAPs, and even satellites, enabling flexible and dynamic deployment. This allows for on-demand coverage extension, rapid deployment in disaster scenarios, and adaptive positioning for optimal signal reflection.

2) Enhanced Channel Quality: Strategically positioning RIS-equipped aerial platforms facilitates the establishment of LoS links, minimizing signal blockage and path loss. This leads to improved channel quality, higher achievable data rates, and increased reliability for communication links between satellites, aerial platforms, and ground users [9].

3) Panoramic Full-Angle Reflection: TRIS is typically limited to reflecting signals within a 180° half-space due to its fixed placement on a vertical surface. Mounting an RIS on an aerial platform, however, can provide 360° panoramic full-angle reflection. This capability enables signal reflection and coverage in all directions, offering greater flexibility and efficiency in serving users spread across a wider area.

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 to be carefully optimized 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. 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 eavesdropper, disrupting any attempts at interception.

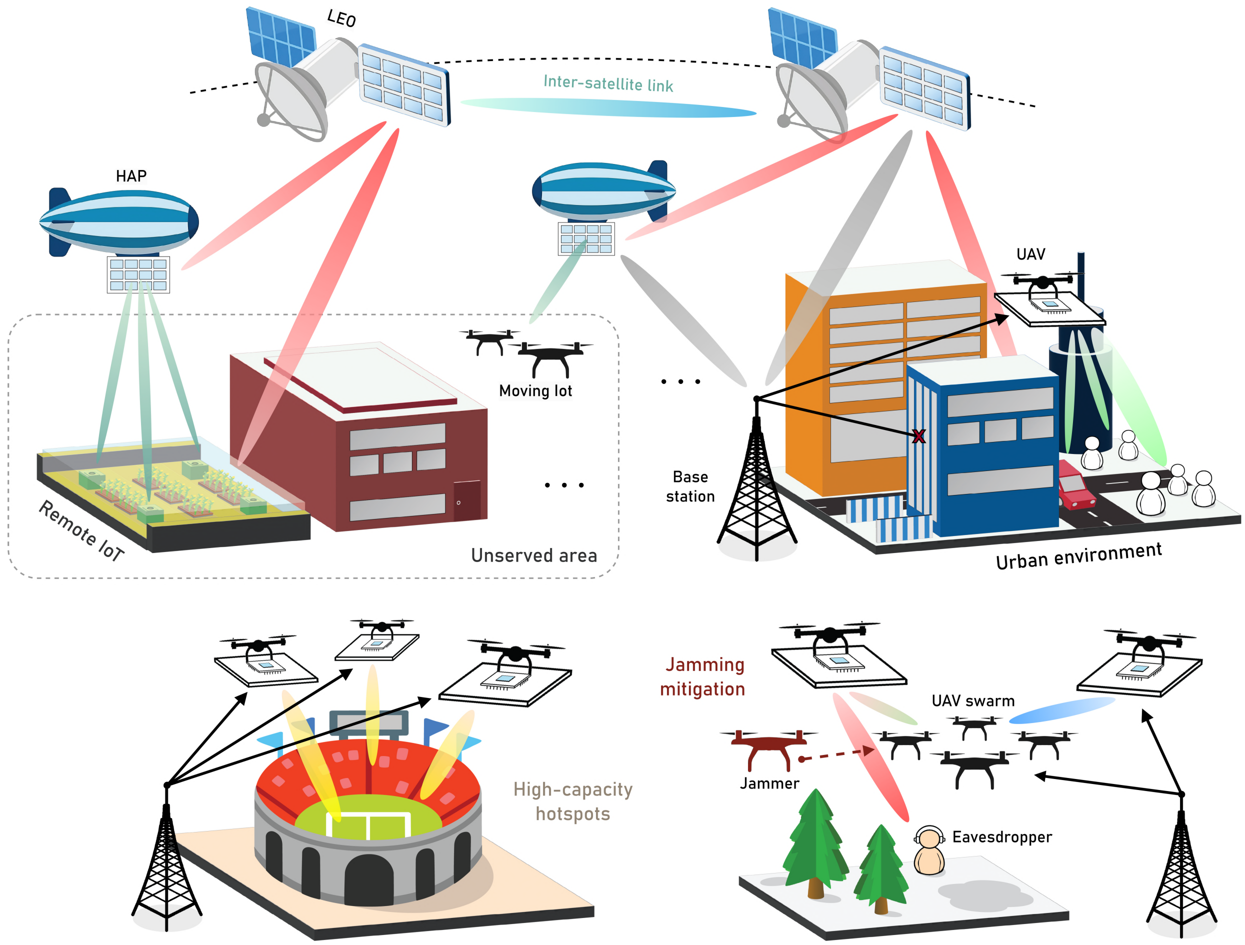

Fig. 1. Illustrative use cases of RIS in aerial NTNs for enhanced connectivity, coverage, and security.

III. Deep Reinforcement Learning for Enhanced RIS-Assisted Aerial NTNs

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 NTNs 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. While conventional optimization techniques, such as convex optimization, have been applied to wireless communication problems, they often struggle to cope with the dynamic and

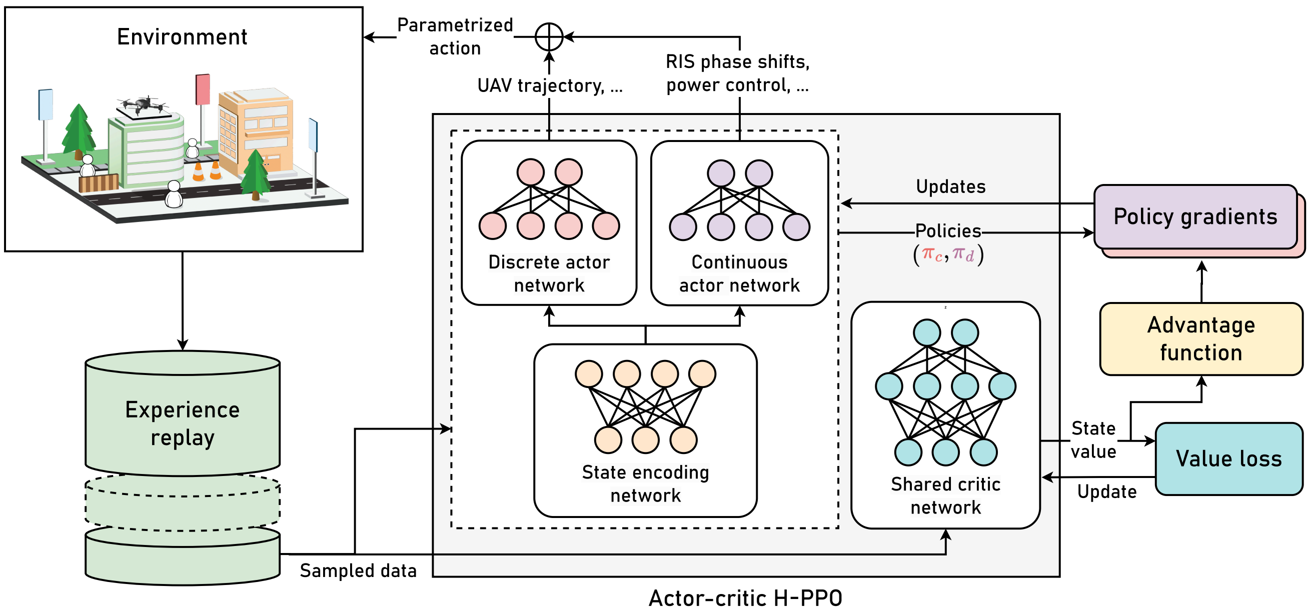

Fig. 2. Architecture of H-PPO for optimization of RIS-assisted aerial NTNs.

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, respectively. In the context of RIS-assisted aerial NTNs, however, we need to control both discrete actions, such as aerial maneuvers, and continuous actions, such as precise adjustments of power control levels and RIS phase shifts.

Applying DQN directly to this scenario would necessitate discretizing the continuous actions, resulting in an impractically large action space that would hinder convergence and make learning inefficient. While DDPG can handle continuous actions, it may not be ideal for the hybrid action spaces in RIS-assisted aerial NTNs. DDPG can exhibit instability when

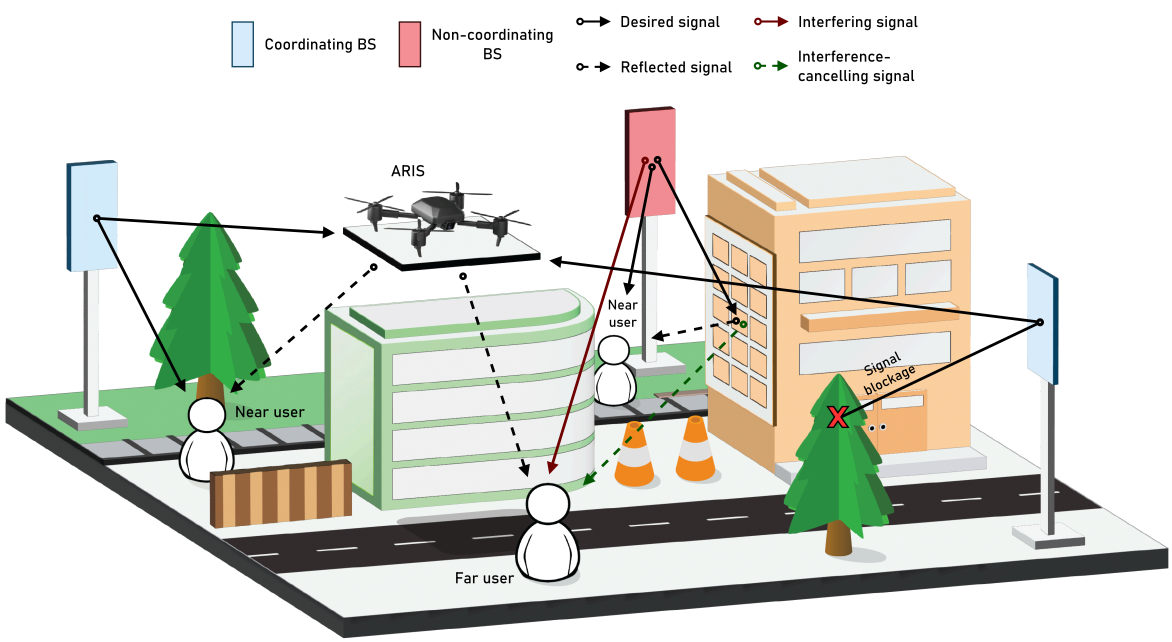

Fig. 3. System model of ARIS-assisted CoMP-NOMA network.

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. These networks, however, present complex optimization challenges, particularly when incorporating ARIS. As discussed earlier, DRL-based optimization provides an efficient and adaptable approach in managing such scenarios.

In this subsection, we investigate the DRL-based optimization of an ARIS-assisted CoMP-NOMA network, demonstrating how H-PPO can optimally allocate resources in order 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 assumption of a substantial propagation distance between them.

All communication channels in the network are modeled 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. 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

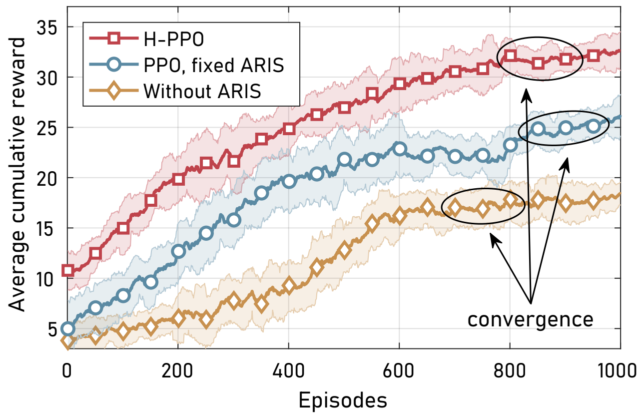

Fig. 4. Average cumulative reward versus the number of training episodes.

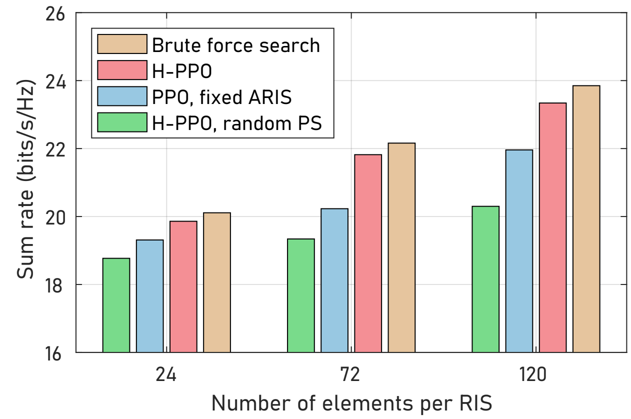

Fig. 5. Network sum rate reward versus the number of elements in both ARIS and TRIS.

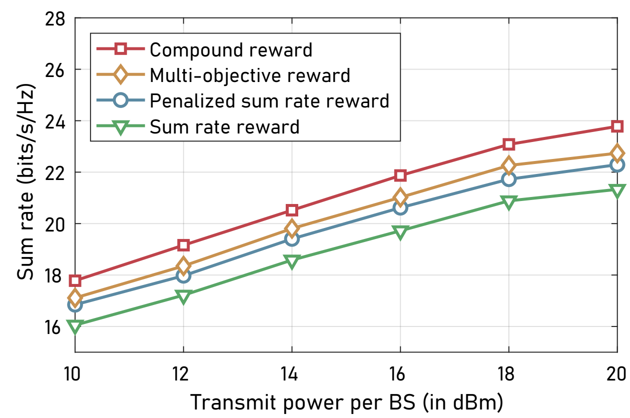

Fig. 6. Network sum rate versus transmit power per BS for different reward functions.

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 150x150 m^2. 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 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. Although a simple sum rate reward might seem intuitive for maximizing sum rate, it can lead to suboptimal performance as the agent may prioritize actions that yield a momentarily high sum rate but violate operational constraints, such as the UAV leaving the designated area or assigning power allocation factors that result in unsuccessful SIC. Incorporating penalties for such violations, as in the penalized sum rate reward, encourages the agent to balance sum rate maximization with constraint adherence. Multi-objective rewards enable an agent to optimize multiple objectives simultaneously, each with a distinct reward function. This method allows exploration of trade-offs between conflicting goals, like achieving high sum rates and maintaining user fairness. In contrast, compound rewards combine multiple reward signals into a single function through weighted sums, simplifying 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, suggesting that holistic reward designs can be more effective.

IV. Challenges and Future Directions

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

A. Algorithmic Challenges

The application of DRL to optimize RIS-assisted aerial NTNs presents several algorithmic challenges that demand further investigation.

1) *Sample Efficiency:* Model-free DRL algorithms, such as DDPG and PPO, often require extensive interactions with the environment, leading to sample inefficiency, especially in complex settings like aerial NTNs. Improving the sample efficiency of DRL algorithms, potentially through techniques like meta-learning, transfer learning, or model-based RL, is crucial for accelerating training and enabling faster adaptation to changing network conditions [12].
2) *Reward Shaping and Exploration:* Designing effective reward functions for aerial NTNs is challenging, requiring a balance between multiple objectives like sum rate maximization, power minimization, and user fairness. Encouraging exploration while maintaining stability is also crucial, particularly in high-dimensional action spaces. Techniques like intrinsic motivation or curiosity-driven exploration can be investigated to enhance DRL agent exploration strategies.
3) *Generalization and Transfer Learning:* DRL policies trained for specific NTN scenarios may not generalize well to environments with different user distributions, channel conditions, or platform configurations. Developing techniques for training DRL agents that can generalize across diverse scenarios or effectively transfer learned knowledge to new environments is crucial for realizing the adaptability and scalability of DRL-powered solutions.

B. Implementation Challenges

Several system-level challenges need to be overcome for a successful ARIS implementation.

1) *Channel Estimation and Modeling:* Accurate channel estimation and modeling are crucial for optimizing RIS configurations in aerial NTNs [13]. The dynamic movement of aerial platforms and the passive nature of RIS necessitate efficient and robust channel estimation protocols, potentially leveraging compressed sensing, machine learning, or novel pilot signaling. Incorporating platform mobility, RIS characteristics, and environmental factors into channel models can further improve prediction accuracy and system performance.
2) *Robustness and Security:* RIS-assisted aerial NTNs must be resilient to uncertainties like imperfect CSI, UAV wobbling, and environmental factors. Designing robust beamforming algorithms and addressing security concerns related to RIS control signaling, communication links, and vulnerabilities to jamming or eavesdropping are paramount. Robust security protocols, encryption techniques, and intrusion detection mechanisms are essential for safeguarding these networks.
3) *Power Consumption and Efficiency:* While RIS elements are passive, the overall power consumption of RIS-assisted aerial NTNs, especially for battery-powered platforms, needs careful consideration. This includes power for propulsion, RIS control, and onboard processing. Optimizing power allocation strategies and incorporating energy-efficient algorithms are essential for maximizing operational time and range.

C. Emerging Directions

Several emerging research directions promise to further enhance ARIS capabilities and broaden its applications.

1) *Edge Computing:* Integrating RIS-assisted aerial NTNs with edge computing can bring computation and storage resources closer to users, enabling low-latency processing and localized data management for applications like real-time video analytics, drone traffic control, and distributed AI [14]. This integration can significantly improve network efficiency and support the growing demands of data-intensive and latency-sensitive applications.
2) *Distributed Learning:* Distributed learning techniques, such as federated learning, can enable collaborative learning across multiple RIS-equipped platforms and ground users without sharing raw data. This approach enhances privacy, improves communication efficiency, and increases scalability. Distributed learning allows aerial platforms to share learned knowledge and adapt to changing network conditions more effectively, especially in large-scale deployments.
3) *ARIS-Assisted ISAC:* Integrating ARIS with ISAC systems offers promising opportunities for enhancing both communication and sensing capabilities [15]. By intelligently controlling RIS phase shifts and aerial trajectories, beamforming can be optimized for both functionalities, maximizing performance and balancing the inherent trade-off between them.
4) *Active ARIS:* Aerial active RIS (AARIS), incorporating amplifiers into reflecting elements, can further enhance aerial NTN coverage, capacity, and link reliability by jointly amplifying and reflecting signals [12]. This potentially overcomes double path loss limitations but introduces new energy efficiency optimization challenges due to the increased power consumption of active elements.

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 remain in terms of scalability, channel estimation, robustness, and security, ongoing research is actively exploring solutions and paving the way for the practical implementation of DRL-powered RIS-assisted aerial NTNs. Emerging directions, including the integration with edge computing, distributed learning, ISAC systems, and active RIS technologies, further broaden the potential of these networks and promise to drive innovation in the rapidly evolving landscape of next-generation wireless communications.

References

[1] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[2] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, 2016.

[3] J. Hu, H. Zhang, L. Song, Z. Han, and H. V. Poor, "Reinforcement learning for a cellular internet of UAVs: Protocol design, trajectory control, and resource management," *IEEE Wirel. Commun.*, vol. 27, no. 1, pp. 116–123, 2020.

[4] S. Basharat, S. A. Hassan, H. Pervaiz, A. Mahmood, Z. Ding, and M. Gidlund, "Reconfigurable intelligent surfaces: Potentials, applications, and challenges for 6G wireless networks," *IEEE Wirel. Commun.*, vol. 28, no. 6, pp. 184–191, 2021.

[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] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 2, pp. 244–251, 2021.

[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, 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] X. Tang, D. Wang, R. Zhang, Z. Chu, and Z. Han, "Jamming mitigation via aerial reconfigurable intelligent surface: Passive beamforming and deployment optimization," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6232–6237, 2021.

[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.