Scientific Reports

https://doi.org/10.1038/s41598-026-66689-6

Article in Press

Continuous control deep reinforcement learning for joint resource optimization in STAR-RIS assisted mmWave-NOMA systems with hardware impairments

Received: 29 April 2026

Accepted: 10 August 2026

Published online: 15 August 2026

Cite this article as: Sobhi-Givi S., Nouri M., Abdollahvand M. *et al.* Continuous control deep reinforcement learning for joint resource optimization in STAR-RIS assisted mmWave-NOMA systems with hardware impairments. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-66689-6

Sima Sobhi-Givi, Mahdi Nouri, Mousa Abdollahvand, Hamid Behroozi, Zhiguo Ding & Zahriladha Zakaria

Continuous Control Deep Reinforcement Learning for Joint Resource Optimization in STAR-RIS Assisted mmWave-NOMA Systems with Hardware Impairments

Sima Sobhi-Givi[1*], Mahdi Nouri[2], Mousa Abdollahvand[1], Hamid Behroozi[2], Zhiguo Ding[3], and Zahriladha Zakaria[4]

[1]Electrical and Computer Engineering Department, University of Mohaghegh Ardabili, Ardabil, Iran

[2]Electrical Engineering Department, Sharif University of Technology, Tehran, Iran

[3]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

[4]Centre for Telecommunication Research and Innovation (CeTRI), Faculty of Electronic and Computer Technology and Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, MALAYSIA

[*] Corresponding Author: Sima Sobhi-Givi (Email: s.sobhi@uma.ac.ir)

Abstract

This paper proposes a continuous-control deep reinforcement learning (DRL) framework for the joint optimization of power allocation, hybrid beamforming, and phase-shift configuration in simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS)-assisted millimeter-wave non-orthogonal multiple access (mmWave-NOMA) networks. The proposed system supports heterogeneous services, including enhanced mobile broadband (eMBB) and ultra-reliable low-latency communication (uRLLC), by maximizing the minimum signal-to-interference-plus-noise ratio (SINR) among users while satisfying quality-of-service requirements. Practical hardware impairments, including imperfect successive interference cancellation (SIC) and in-phase and quadrature imbalance (IQI), are explicitly incorporated into the system model to capture residual interference and transceiver distortions. A continuous-control DRL framework based on Softmax Deep Double Deterministic Policy Gradients (SD3) is developed to jointly optimize the network resources under these non-ideal conditions. Extensive simulation results demonstrate that the proposed STAR-RIS-assisted SD3 framework consistently outperforms conventional RIS-based schemes and benchmark DRL algorithms in terms of achievable sum rate and convergence performance. Furthermore, robustness evaluations under imperfect channel state information (CSI), varying hardware impairment levels, and different Rician fading conditions show that the proposed approach maintains superior performance despite channel uncertainty and practical transceiver impairments, highlighting its effectiveness for next-generation mmWave communication systems.

Index Terms

Network slicing, non-orthogonal multiple access (NOMA), STAR-RIS, hybrid beamforming, deep reinforcement learning (DRL).

I. Introduction

A. Background and Motivation

THE evolution toward sixth-generation (6G) wireless networks is driven by the need to simultaneously support heterogeneous services with stringent and often conflicting requirements, such as ultra-high data rates, ultra-low latency, high reliability, and massive connectivity. While 5G has introduced several enabling technologies, it remains limited in addressing these diverse demands in a scalable and energy-efficient manner. Consequently, 6G research has increasingly focused on the joint design of intelligent radio environments, advanced multiple access techniques, and flexible network architectures.

Reconfigurable intelligent surfaces (RISs) have emerged as a key enabling technology for beyond-5G and 6G wireless networks owing to their ability to intelligently manipulate the wireless propagation environment while improving spectral efficiency, coverage, and energy efficiency [1]. Recent surveys have further highlighted their significant potential in future communication, sensing, and green Internet-of-Everything applications [2]. However, conventional RIS architectures are limited to reflection-only operation, meaning that incident signals can only be redirected toward one side of the surface. This half-space coverage restricts deployment flexibility and limits the ability to simultaneously serve users located on opposite sides of the RIS. To overcome this limitation, simultaneously transmitting and reflecting RIS (STAR-RIS) has recently been proposed. By jointly controlling the transmission and reflection coefficients of each RIS element, STAR-RIS enables simultaneous signal transmission and reflection, thereby providing full-space coverage and significantly improving coverage flexibility, user connectivity, and resource utilization. These advantages make STAR-RIS particularly attractive for heterogeneous wireless networks, where users with different quality-of-service requirements may be distributed on both sides of the intelligent surface [3]–[6].

In parallel, the rapid growth of data traffic and spectrum scarcity in 6G networks have propelled millimeter-wave (mmWave) communications to the forefront due to their abundant bandwidth and ability to support large antenna arrays [7], [8]. When combined with massive multiple-input multiple-output (MIMO) systems, mmWave technology significantly enhances spectral efficiency (SE) through high beamforming gains. Hybrid beamforming, which integrates analog and digital processing, is particularly important in mmWave systems to balance performance and hardware complexity [9], [10]. However, the highly directional nature of mmWave transmissions leads to challenges in user scheduling, coverage continuity, and efficient resource utilization, especially in dense and heterogeneous service scenarios.

To address the inefficiencies of conventional orthogonal multiple access (OMA) schemes—such as time-division and frequency-division multiple access—that limit one user per resource block, power-domain non-orthogonal multiple access (NOMA) has been proposed as a key multiple access technique for future networks [11]–[13]. By enabling multiple users to share the same time–frequency resources and employing successive interference cancellation (SIC) at the receiver, NOMA improves spectral efficiency while supporting fairness and diverse quality-of-service (QoS) requirements. Importantly, the spatially correlated channels created by narrow mmWave beams make mmWave systems particularly well-suited for NOMA transmission [14]–[16]. Nevertheless, the performance of mmWave–NOMA systems critically depends on effective beamforming design, power allocation, and interference management.

Beyond the physical layer, network slicing has been recognized as a fundamental architectural enabler for 6G, allowing multiple virtual networks to coexist on a shared physical infrastructure [17]. Each slice can be customized to meet the specific requirements of different services, such as enhanced mobile broadband (eMBB) and ultra-reliable low-latency communication (uRLLC) [18], [19]. While eMBB prioritizes high throughput, uRLLC demands stringent latency and reliability guarantees, leading to inherent tradeoffs in resource allocation [20], [21]. Traditional network slicing approaches rely on orthogonal resource partitioning, which results in inefficient spectrum utilization. To overcome this limitation, recent studies have explored physical-layer-aware slicing strategies, including heterogeneous OMA (Het-OMA) and heterogeneous NOMA (Het-NOMA), where multiple services share the same physical resources [22]–[24]. In particular, Het-NOMA exploits SIC to enable service coexistence, achieving notable performance gains by leveraging the robustness of uRLLC decoding in the presence of eMBB interference.

Motivated by these observations, the integration of mmWave communications, advanced RIS architectures, NOMA-based multiple access, and network slicing emerges as a promising paradigm for addressing the fundamental challenges of 6G networks. Such a joint design has the potential to simultaneously enhance spectral efficiency, coverage flexibility, and service differentiation while maintaining energy efficiency and scalability. Meanwhile, the joint optimization of beamforming, RIS configuration, and resource allocation results in highly coupled and non-convex optimization problems, making conventional optimization techniques computationally expensive for practical deployment. Recent studies have demonstrated that learning-based methods provide an effective alternative for adaptive optimization and interference management in RIS-assisted wireless networks [25]. Motivated by these advances, this work investigates a STAR-RIS-assisted mmWave-NOMA network for heterogeneous sliced services and develops a continuous-control deep reinforcement learning framework for joint resource optimization.

B. Related Work

In [26], a joint optimization problem for STAR-RIS-NOMA systems was investigated, with the goal of maximizing the achievable sum-rate. This was achieved by iteratively optimizing the decoding order, active beamforming, passive transmission and reflection beamformings, along with the power allocation coefficients. In [26], this challenging optimization problem was divided into separate sub-problems, which are solved in an alternating manner. As a result, their proposed approach is not optimal and does not offer a comprehensive solution.

In [27] the deployment of an RIS to enhance communication between different services and a common base station (BS) was investigated. It evaluates various radio access network (RAN) slicing strategies, including Het-OMA and Het-NOMA.

The study in [28] addressed STAR-RIS channel models by proposing distinct models for near-field and far-field regions. Additionally, it derives closed-form expressions for the channel gains experienced by users receiving both transmission and reflection signals. In [29], the authors explored the fundamental coverage characteristics of STAR-RIS-assisted two-user communication networks. The study formulated problems aimed at maximizing the sum coverage range for both NOMA and OMA systems. Furthermore, [30] addressed power consumption minimization in STAR-RIS-assisted unicast and multicast systems, where both active and passive beamforming were optimized together for various STAR-RIS operating protocols.

In [31], a general optimization framework for rate splitting multiple access (RSMA) in beyond diagonal (BD) RIS-assisted uRLLC systems was proposed. This framework provides suboptimal solutions for a wide range of optimization problems where the objectives and constraints are linear functions of user rates and energy efficiency (EE). The study reveals that RSMA and RIS are particularly beneficial in overloaded systems, where the number of users per cell exceeds the number of base station (BS) antennas. As previously mentioned, one of the key objectives of 6G is to significantly enhance SE and EE, which are particularly crucial in uRLLC systems. RIS have emerged as a promising technology to achieve these targets. In [32], a NOMA-based scheme for a multi-cell multiple-input single-output (MISO) RIS-assisted broadcast channel (BC) was proposed. The study demonstrated that RIS could substantially increase both the minimum rate and the EE of the network, highlighting its potential in optimizing network performance. Similarly, [33] showed that RIS could significantly enhance the SE and EE of a multi-cell multiple-input multiple-output (MIMO) BC, even when the transceivers experience in-phase/quadrature (I/Q) imbalance. This indicates the robustness of RIS in improving communication system performance under various conditions.

The application of RIS technology to solve the resource allocation problem was explored in [34]. In this study, the resource allocation challenge is formulated into two separate optimization problems. For the eMBB allocation, the authors jointly optimize the resource block allocation, power distribution, and the RIS phase shift matrix, proposing a two-stage alternating

TABLE I: SUMMARY OF PROPOSED METHOD AND RELATED WORKS

	Objective	Network Slicing	NOMA	STAR-RIS	Power Allocation	Beamforming	max-min Fairness	Solution
Proposed	Designing active and passive beamforming matrix, power allocation, Max-min SINR	✓	✓	✓	✓	✓	✓	DDPG, TD3, SD3
[26]	Designing active and passive beamforming matrix, power allocation, Maximizing sum-rate	×	✓	✓	✓	✓	×	Iterative Algorithm
[27]	Enhancing communication between different services	✓	✓	×	×	×	×	–
[29]	Resource allocation, Maximizing sum-coverage	×	✓	✓	×	×	×	Search-based Algorithm
[30]	Designing active and passive beamforming, Minimizing power consumption	×	×	✓	×	✓	×	Iterative Algorithm
[31]	Designing beamforming, Maximizing SE and EE	✓	×	×	×	✓	×	Iterative Algorithm
[32]	Power allocation, Maximizing minimum weighted rate and EE	×	✓	×	✓	×	✓	Improper Gaussian Signaling Algorithm
[34]	Designing passive beamforming, power and RB allocation, Maximizing capacity and minimizing rate loss	✓	×	×	✓	✓	×	Iterative Algorithm
[36]	Designing passive beamforming, positioning UAV, Minimizing total decoding error rate	✓	×	×	×	✓	×	Direct Search Method

iterative algorithm to maximize the total throughput capacity for eMBB users. On the other hand, for the uRLLC allocation, the focus shifts to maximizing the uRLLC packet reception rate while minimizing the rate loss of eMBB users, ensuring high-quality service for both uRLLC and eMBB. A heuristic uRLLC allocation algorithm based on a pre-configured RIS was introduced to meet these objectives.

The works in [35]–[38] concentrated on single-service applications, either for eMBB or uRLLC. In [35], it was demonstrated that RIS-assisted MIMO systems could achieve performance comparable to large-scale MIMO systems without RIS, but with a significant reduction in the number of active antennas and RF chains. [36] focused on UAV-assisted short data uRLLC scenarios, where the passive beamforming capabilities of RIS, combined with a direct search method, were used to optimize the UAV's location and block length in the theoretical model, resulting in a reduction in total decoding BER. In [37], an authorization-free access scheme was proposed, leveraging RIS to enhance the reliability of uRLLC services. Additionally, [38] presented a suboptimal iterative algorithm that employs a novel iterative rank minimization method to optimize all system variables in each iteration, simplifying uRLLC services and yielding substantial performance improvements. Moving beyond single-service cases, [39] investigated the coexistence of both eMBB and uRLLC services in RIS-assisted cellular networks.

The authors in [40] presented a multi-sector beyond diagonal RIS (BD-RIS) model, offered a detailed channel model that includes antenna gain and demonstrated that their model can encompass configurations such as STAR-RIS. The paper derives a scaling law for received power, showing a positive correlation with the number of sectors in a BD-RIS-assisted SU-SISO system. Additionally, it examines the application of multi-sector BD-RIS in MU-MISO systems to maximize sum-rate under different amplitude and phase shift settings.

C. Contributions

Considering the existing literature, a noticeable research gap exists integrating hybrid beamforming, STAR-RIS phase shift matrix design, and power allocation within networks that simultaneously accommodate uRLLC and eMBB slices. This paper addresses this gap by incorporating Het-NOMA into network slicing systems. The aim of this paper is to achieve max-min fairness using the SINR metric, for which deep reinforcement learning (DRL) are employed specifically the softmax deep double deterministic policy gradients (SD3) algorithm. Therefore, the main contributions of this paper are:

- A comprehensive examination of hybrid mmWave-NOMA networks, focusing on the coexistence of uRLLC and eMBB services, considering the practical challenges posed by imperfect SIC and IQI impairments.
- Introduction of STAR-RIS deployment in scenarios with obstructed direct transmission paths, providing full-space coverage that traditional RIS models cannot achieve, thus enhancing the network's capability to serve both uRLLC and eMBB slices.
- Design of a general framework for analyzing and maximizing the minimum SINR to ensure max-min fairness in mmWave-NOMA networks, particularly under the challenging conditions introduced by imperfect SIC and IQI.
- Introduction of DRL-based methods for the joint optimization of power allocation, beamforming design, and STAR-RIS phase shift configuration, leveraging the SD3 algorithm to achieve the trade-offs between network efficiency and fairness, effectively.
- Evaluation of the sum-rate performance of user clusters using various DRL methods, demonstrating the superior performance of the SD3 algorithm. The study also shows that increasing transmit power improves the sum-rate of clusters, and it compares the effectiveness of the proposed STAR-RIS system with traditional RIS systems, highlighting the advantages of the former in enhancing overall network performance, even in the presence of hardware impairments.

Table I summarizes the characteristics of the related works and the proposed network.

The rest of this paper is organized as follows. Section II presents the system model of the proposed mmWave-NOMA transmission network consisting of two slices and STAR-RIS. In Sections III, the proposed optimization problem and the solving solution are explained. Numerical results are provided in Section IV, and finally, Section V concludes the paper.

II. System Description

This paper investigates the mmWave-NOMA system configuration, comprising a BS equipped with N antennas and serving K users, each equipped with a single antenna. BS and k-th user are located at the $\mathbf{b} = [x_b, y_b, z_b]$ and $\mathbf{u}_k = [x_k, y_k, z_k]$, respectively. Due to obstructive elements such as tall buildings, the direct communication link between the BS and users is assumed to be blocked. In such scenarios, the STAR-RIS, equipped with N_{SR} elements, acts as a relay to enhance communication between the BS and users, which is located at the $\mathbf{r} = [x_r, y_r, z_r]$. To manage the system effectively, M ($M < N$) radio frequency (RF) chains are deployed at the BS, facilitating the transmission of N_s independent data streams to the user terminals. To optimize the multiplexing gain, we assume $N_s = M$, aligning the number of data streams with the RF chains. Initially, digital beamforming (DBF) precodes the N_s data streams using the baseband matrix $\mathbf{D} \in \mathbb{C}^{M\times N_s}$. After undergoing RF processing, the signal in the digital domain traverses N phase shifters with a constant modulus for analog beamforming (ABF), utilizing the beamforming matrix $\mathbf{A} \in \mathbb{C}^{N\times M}$. Consequently, the hybrid beamforming matrix, denoted as $\mathbf{W} = \mathbf{AD} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M]$, is formed, with each column normalized, i.e., $\|\mathbf{w}_m\| = 1,\ 1 \leq m \leq M$ [41]. Furthermore, to simplify expressions of the phase shift matrix of STAR-RIS, the transmission ($p = t$) or reflection ($p = r$) beamforming vector is denoted as $\mathbf{v}_p = \left[\sqrt{\beta_1^p}e^{j\theta_1^p}, \sqrt{\beta_2^p}e^{j\theta_2^p}, \ldots, \sqrt{\beta_{N_{SR}}^p}e^{j\theta_{N_{SR}}^p}\right]^H$, and $\mathbf{\Theta}_p = \mathrm{diag}(\mathbf{v}_p)$ represents the corresponding diagonal beamforming matrix. Moreover, we refer to the set of variables to the transmission and reflection coefficients as [42]:

$$\mathbb{R}_{\beta,\theta} =\{\beta_n^t, \beta_n^r,\, \theta_n^t, \theta_n^r, |\beta_n^t, \beta_n^r \in [0,1]; \beta_n^t + \beta_n^r = 1; \theta_n^t, \theta_n^r \in [0, 2\pi]\}. \quad (1)$$

In NOMA transmission, K users are grouped into M clusters, each representing an independent data stream. Within each cluster, users can employ NOMA and SIC techniques to manage intra-cluster interference, while inter-cluster interference is addressed through beamforming. With M RF chains supporting a maximum of M data streams, ensuring at least one user in each cluster is essential to prevent underutilizing RF resources. In the proposed network, the BS supports two groups of users requiring uRLLC and eMBB services. The system operates in a downlink (DL) mode, with the BS transmitting to the users. The overall system diagram is depicted in Fig. 1. To ensure low-latency services, each time slot, with a duration Δ, is partitioned into S mini-slots, each lasting δ. This division aids in managing latency sensitive services efficiently. For eMBB traffic, a single time-slot is sufficient. Conversely, for a uRLLC device a single mini-slot is used to leverage frequency diversity and meet reliability criteria [23]. Given the stringent latency requirements, each uRLLC packet must be decoded within a mini-slot's duration and cannot span multiple mini-slots [43].

The Het-NOMA slicing of the radio resources for eMBB and uRLLC users is illustrated in Fig. 1, where the RBs can be shared between the two services. In summary, each NOMA group consists of one eMBB user and one uRLLC user. The design of hybrid beamforming, phase shift matrix, and power allocation are achieved through DRL algorithms.

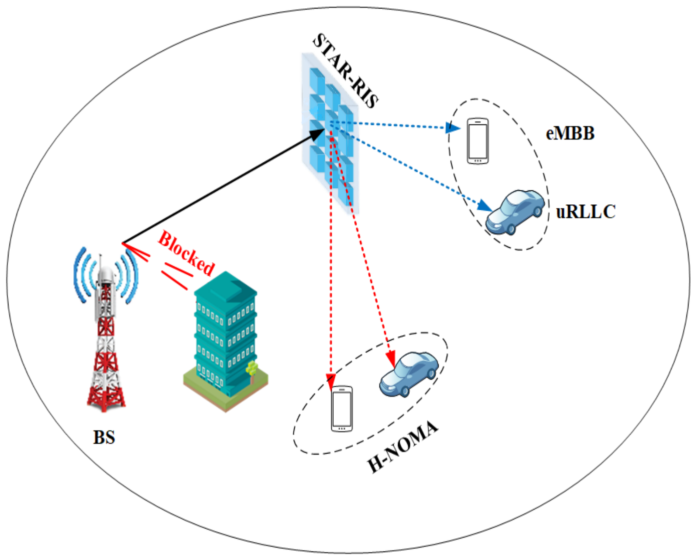


Fig. 1: System model enabling coexisting eMBB and uRLLC services with STAR-RIS assistance in 5G/B5G networks

A. Channel Model

1) Perfect CSI Model: The channel between the BS and STAR-RIS, $\mathbf{G} \in \mathbb{C}^{N_{SR} \times N}$, is assumed to be LoS and can be expressed as [44]:

$$\mathbf{G} = \sqrt{\frac{\rho_0}{||\mathbf{b}-\mathbf{r}||^2}}\bar{\mathbf{G}}, \tag{2}$$

where ρ_0 is the path loss at the reference distance of one meter. $||.||^2$ represents the Euclidean norm operation. $\bar{\mathbf{G}}$ shows the deterministic LoS component and defined as:

$$\bar{\mathbf{G}} = [1, e^{-j\frac{2\pi d}{\lambda}\cos\phi}, \cdots, e^{-j\frac{2\pi d}{\lambda}(N_{SR}-1)\cos\phi}]^T, \tag{3}$$

where $\cos\phi = \frac{x_b - x_r}{||\mathbf{b}-\mathbf{r}||}$ is the cosine of the angle of arrival (AoA) from the BS to the STAR-RIS. Since mmWave channels generally exhibit a dominant line-of-sight component together with several scattered paths, Rician fading has been widely adopted for RIS-assisted wireless systems to accurately characterize channel propagation [45]. Also, the channel between the STAR-RIS and users, $\mathbf{h}_i \in \mathbb{C}^{N_{SR} \times 1}$, can be displayed as:

$$\mathbf{h}_k = \sqrt{\frac{\rho_0}{||\mathbf{r}-\mathbf{u}_k||^{\alpha}}}\left(\sqrt{\frac{\beta}{1+\beta}}\bar{\mathbf{h}}_k + \sqrt{\frac{1}{1+\beta}}\hat{\mathbf{h}}_k\right), \tag{4}$$

where α represents the path loss exponent of the RIS to user link, β denotes the Rician factor. The deterministic LoS component is represented by $\bar{\mathbf{h}}_k$ and defined similarly to (3), while $\hat{\mathbf{h}}_k$ characterizes the random Rayleigh distributed non-LoS (NLoS) component. With the channel models, the effective power gain between the BS and k-th user assisted by the STAR-RIS, $\mathbf{H}_k \in \mathbb{C}^{1 \times N}$, is given by

$$\mathbf{H}_k = \left|\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G} \mathbf{w}_m\right|^2, \quad k \in \{uRLLC, eMBB\}, \tag{5}$$

where $(.)^H$ is the Hermitian.

B. Imperfect CSI Model

In practical STAR-RIS-assisted mmWave systems, acquiring perfect CSI is challenging due to the large number of reflecting/transmitting elements and the cascaded channel structure. Therefore, the impact of channel estimation errors is considered by modeling the actual channel as

$$\mathbf{h}_i = \hat{\mathbf{h}}_i + \Delta\mathbf{h}_i, \tag{6}$$

where $\hat{\mathbf{h}}_i$ denotes the estimated channel and $\Delta\mathbf{h}_i$ represents the channel estimation error. The estimation error is assumed to follow a circularly symmetric complex Gaussian distribution as $\Delta\mathbf{h}_i \sim \mathcal{CN}(0, \sigma_e^2\mathbf{I})$, where σ_e^2 controls the CSI uncertainty level. Similarly, the estimated BS–STAR-RIS channel and STAR-RIS–user channels respectively can be represented as

$$\mathbf{G} = \hat{\mathbf{G}} + \Delta\mathbf{G}, \tag{7}$$

and

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \Delta\mathbf{h}_k, \tag{8}$$

These channel estimation errors introduce a mismatch between the actual propagation environment and the available CSI, which affects the accuracy of the obtained beamforming and STAR-RIS transmission coefficients.

C. Signal Model

1) Perfect system: In a NOMA transmission scenario, the BS simultaneously transmits signals to K users by superimposing their individual symbols into a single transmitted signal x. This signal is formulated as:

$$x = \sum_{k=1}^{K} \sqrt{c_k P} x_k, \ \ k \in \{uRLLC, eMBB\} \tag{9}$$

where c_k represents the power allocation factor for the k-th user, x_k is the symbol intended for the k-th user, and P denotes the total transmission power of the BS. The symbols are normalized such that $E[|x_k|^2] = 1$, and the power allocation factors sum to unity, i.e., $\sum_{k=1}^{K} c_k = 1$.

The users are arranged in ascending order based on their channel gains, such that the first user has the weakest channel and the last user has the strongest channel. Accordingly, higher power coefficients are allocated to users with weaker channel conditions, while users with stronger channels are assigned lower power. This ordering is essential for the SIC process. During SIC, the k-th user first decodes and cancels the signals intended for users 1 through $k-1$(i.e., users with weaker channels and higher power allocation), and then decodes its own signal. The signals of users $k+1$ through K are treated as residual interference. Assuming perfect SIC at the k-th user, the received signal (y_k) for this user can be expressed as:

$$y_k = \underbrace{\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G} \mathbf{w}_m \sqrt{c_k P} x_k}_{\text{Desired signal term}} + \underbrace{\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G} \mathbf{w}_m \sum_{l=k+1}^{K} \sqrt{c_l P} x_l}_{\text{User-interference term}} + n_k, \tag{10}$$

where n_k denotes the complex additive white Gaussian noise (AWGN) sample with zero mean and variance σ_k^2 at the k-th user, $n_k \sim \mathcal{CN}(0, \sigma_k^2)$.

2) Imperfect system: In this scenario, it is assumed that the strong user cannot perfectly reconstruct and cancel the decoded signals during the SIC process. Consequently, a fraction of the interference associated with the previously decoded users remains after cancellation. This residual interference is modeled through an imperfect SIC coefficient, which captures the effects of channel estimation errors, hardware impairments, and practical receiver limitations.

In this setup, the transmit signal x remains as defined in (9), where the BS transmits a superimposed signal composed of the individual symbols for all K users, i.e.,

$$\hat{x} = x, \tag{11}$$

After performing SIC, the decoded users' signals cannot be completely removed from the received signal. Therefore, a residual interference term proportional to the imperfect SIC coefficient remains, while the signals of users that are not decoded continue to act as ordinary multi-user interference. Accordingly, the received signal at user k under imperfect SIC can be expressed as:

$$\hat{y}_k = \underbrace{\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G} \mathbf{w}_m \sqrt{c_k P} x_k}_{\text{Desired signal term}} + \xi \underbrace{\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G} \mathbf{w}_m \sum_{l=1}^{k-1} \sqrt{c_l P} x_l}_{\text{Residual SIC}} + \underbrace{\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G} \mathbf{w}_m \sum_{l=k+1}^{K} \sqrt{c_l P} x_l}_{\text{User-interference term}} n_k, \tag{12}$$

where $0 \leq \xi \leq 1$ denotes the imperfect SIC coefficient. The value $\xi = 0$ corresponds to perfect SIC, while larger values indicate stronger residual interference due to incomplete cancellation of the previously decoded users' signals.

3) IQI impairment: In practical communication systems, hardware imperfections may arise from components such as the phase shifter, local oscillator (LO), and IQ mixer, resulting in amplitude and phase mismatches between the in-phase (I) and quadrature (Q) branches. These impairments, commonly referred to as in-phase and quadrature imbalance (IQI), distort both the transmitted and received signals, thereby degrading the overall system performance. Specifically, phase imbalance causes the phase difference between the I and Q components to deviate from the ideal 90-degree, while amplitude imbalance results from unequal gains of the I and Q branches. Both impairments introduce an image signal that cannot be completely eliminated, leading to performance degradation.

To model the transmitter IQI, the equivalent time-domain baseband signal, $\bar{x}$, is expressed as

$$\bar{x} = \mu_t x + \nu_t^* x^* = \mu_t \sum_{k=1}^{K} \sqrt{c_k P} x_k + \nu_t^* \sum_{k=1}^{K} \sqrt{c_k P} x_k^*, \tag{13}$$

where x^* denotes the complex conjugate of the transmitted signal, while μ_t and ν_t represent the transmitter IQI coefficients. Following [46], these coefficients are given by

$$\mu_t \triangleq \frac{1 + g_t e^{j\phi_t}}{2}, \tag{14}$$

$$\nu_t \triangleq \frac{1 - g_t e^{-j\phi_t}}{2}, \tag{15}$$

where g_t and ϕ_t denote the transmitter amplitude and phase mismatches, respectively.

The IQI-impaired transmitted signal propagates through the STAR-RIS-assisted channel before being affected by the receiver IQI. Accordingly, the received baseband signal prior to receiver IQI is expressed as

$$y_k = \mathbf{h}_k^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m \bar{x} + n_k = \mathbf{h}_k^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m \left(\mu_t \sum_{l=1}^{K} \sqrt{c_l P} x_l + \nu_t^* \sum_{l=1}^{K} \sqrt{c_l P} x_l^* \right) + n_k \tag{16}$$

Applying the receiver IQI yields

$$\begin{aligned} \bar{y}_k &= \mu_r y_k + \nu_r y_k^* \\ &= \underbrace{\mu_r \left(\mathbf{h}_k^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m \left(\mu_t \sum_{l=1}^{K} \sqrt{c_l P} x_l + \nu_t^* \sum_{l=1}^{K} \sqrt{c_l P} x_k^* \right) + n_k \right)}_{\text{Direct signal component}} + \underbrace{\nu_r \left(\mathbf{h}_k^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m \left(\mu_t \sum_{l=1}^{K} \sqrt{c_l P} x_l + \nu_t^* \sum_{l=1}^{K} \sqrt{c_l P} x_l^* \right) + n_k \right)^*}_{\text{Image interference component}}, \end{aligned} \tag{17}$$

where y^* denotes the mirror signal introduced by the IQI. The coefficients μ_r and ν_r denote the receiver IQI coefficients, and $(\cdot)^*$ represents the complex conjugate operation. The receiver IQI coefficients are defined as

$$\mu_r \triangleq \frac{1 + g_r e^{j\phi_r}}{2}, \tag{18}$$

$$\nu_r \triangleq \frac{1 - g_r e^{-j\phi_r}}{2}, \tag{19}$$

where g_r and ϕ_r are the receiver amplitude and phase mismatches, respectively, with $g_r \sim \mathcal{N}(0, \sigma_{g_r}^2)$ and $\phi_r \sim \mathcal{N}(0, \sigma_{\phi_r}^2)$.

Finally, the image rejection ratios (IRRs) at the transmitter and receiver are respectively defined as

$$IRR_T \triangleq \frac{|\nu_t|^2}{|\mu_t|^2}, \tag{20}$$

$$IRR_R \triangleq \frac{|\nu_r|^2}{|\mu_r|^2}. \tag{21}$$

D. Signal to Interference and Noise Ratio (SINR)

In this subsection, we derive the signal-to-interference-plus-noise ratio (SINR) expressions for the proposed STAR-RIS-assisted mmWave Het-NOMA network under different operating conditions. Without loss of generality, we consider a single NOMA cluster consisting of one eMBB user and one uRLLC user to facilitate the analytical derivation, while in the simulations 3 NOMA clusters were investigated. The derived SINR expressions can be readily extended to multiple NOMA clusters by applying the same analysis independently to each cluster.

1) Perfect system: As previously discussed, the BS implements Het-NOMA to serve both uRLLC and eMBB users. In Het-NOMA, SIC is utilized at the receiver, enabling the strong user to mitigate interference from weaker users. uRLLC traffic can occur at any mini-slot within a time slot t and requires immediate processing. Each uRLLC traffic must be completed within a mini-slot to meet its latency and reliability requirements. Consequently, the uRLLC user treats the eMBB signal as noise and focuses on decoding its own signal. Therefore, the SINR for the uRLLC user at mini-slot s of time-slot t can be expressed as:

$$\gamma_u^{s,t} = \frac{\left|\mathbf{h}_u^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m\right|^2 c_u P}{\sigma^2 + \left|\mathbf{h}_u^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m\right|^2 c_e P}, \tag{22}$$

where c_u and c_e represent the power gains from the BS to the uRLLC and eMBB users, respectively. σ^2 denotes the noise power. Although the proposed framework optimizes the physical-layer performance through the max-min SINR criterion, packet-level metrics such as packet error rate and latency are not explicitly modeled. In practical uRLLC systems, the achievable rate under finite blocklength transmission can be approximated as

$$R \approx \log_2(1+\gamma) - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon), \tag{23}$$

where n is the blocklength, V denotes the channel dispersion, and ϵ is the target packet error probability. This relationship indicates that improving the received SINR directly enhances the achievable reliability under finite-blocklength communications.

The eMBB user initially decodes the signal of the uRLLC user and applies the SIC technique to remove it. Hence, the SINR of the eMBB user at time-slot t (γ_e^t) can be obtained as:

$$\gamma_e^t = \frac{\left|\mathbf{h}_e^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m\right|^2 c_e P}{\sigma^2}. \tag{24}$$

2) Imperfect system: This section considers the case with imperfect SIC. Despite the imperfections, the SINR for the uRLLC user ($\hat{\gamma}_u^{s,t}$) remains the same as in the case of perfect SIC, and is calculated as:

$$\hat{\gamma}_u^{s,t} = \gamma_u^{s,t}. \tag{25}$$

For the eMBB user, however, the SINR at time slot t ($\hat{\gamma}_e^t$) is affected by the imperfect SIC process. It is defined as:

$$\hat{\gamma}_e^t = \frac{\left|\mathbf{h}_e^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m\right|^2 c_e P}{\sigma^2 + \xi \left|\mathbf{h}_e^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m\right|^2 c_u P}. \tag{26}$$

In this equation, the SINR of the eMBB user is influenced by the residual interference from the uRLLC user due to imperfect SIC. The term ξ represents the level of imperfection in the SIC process, meaning that some interference from the uRLLC user remains and affects the eMBB user's signal. The numerator reflects the desired signal strength for the eMBB user, while the denominator accounts for both the noise and the residual interference from the uRLLC user. This model acknowledges that in practical scenarios, perfect cancellation of interference is rarely achieved, and the residual interference can impact the performance of eMBB users.

3) IQI impairment: Here, IQI impairment is taken into account. The SINR expression shows how the user's signal strength is affected by IQI, leading to potential interference from other users and noise. Therefore, the SINR for the uRLLC user, considering IQI impairment at mini-slot s of time-slot t ($\bar{\gamma}_u^{s,t}$), is given by:

$$\bar{\gamma}_u^{s,t} = \frac{(|A_u|^2 + |B_u|^2) c_u P}{(|A_u|^2 + |B_u|^2) c_e P + (|\mu_r|^2 + |\nu_r|^2)\sigma_n^2}, \tag{27}$$

where c_u and c_e are the NOMA power allocation coefficients for the uRLLC and eMBB users, respectively. The terms $A_u = \mu_r \mu_t \mathbf{h}_u^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m + \nu_r \nu_t \mathbf{w}_m^H \mathbf{G}^H \mathbf{\Theta}^H \mathbf{h}_u$ and $B_u = \mu_r \nu_t^* \mathbf{h}_u^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m + \nu_r \mu_t^* \mathbf{w}_m^H \mathbf{G}^H \mathbf{\Theta}^H \mathbf{h}_u$ represent the desired and image components of the IQI-distorted effective channel, respectively.

Similarly, after decoding and removing the uRLLC signal through SIC, the eMBB user experiences only the IQI-induced distortion and receiver noise. Therefore, assuming perfect SIC, the SINR of the eMBB user at time-slot t, denoted by $\bar{\gamma}_e^t$, is expressed as

$$\bar{\gamma}_e^t = \frac{(|A_e|^2 + |B_e|^2) c_e P}{(|\mu_r|^2 + |\nu_r|^2)\sigma_n^2}. \tag{28}$$

where $A_e = \mu_r \mu_t \mathbf{h}_e^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m + \nu_r \nu_t \mathbf{w}_m^H \mathbf{G}^H \mathbf{\Theta}^H \mathbf{h}_e$ and $B_e = \mu_r \nu_t^* \mathbf{h}_e^H \mathbf{\Theta} \mathbf{G} \mathbf{w}_m + \nu_r \mu_t^* \mathbf{w}_m^H \mathbf{G}^H \mathbf{\Theta}^H \mathbf{h}_e$.

The above expressions show that IQI affects the communication performance by generating an image interference component, which changes the effective channel gain and increases the received distortion. Consequently, both uRLLC and eMBB users experience SINR degradation due to the transmitter and receiver IQ imbalance.

III. Proposed Resource Allocation

In this paper, we present a general framework to analyze the max-min fairness, using performance index based on SINR. In particular, the aim of the paper is on a hybrid mmWave-NOMA network, where the joint design of phase shift matrix at the STAR-RIS, hybrid beamforming at the BS, and power allocation are considered. We assume that the BS has perfect CSI of all users. Users initiate connection requests to the BS, and based on the requests, the BS forms NOMA clusters and designs hybrid beamforming, phase shift matrix, and power allocation for each cluster. This is done by solving the defined optimization problem outlined in the following.

A. Problem Formulation

Our goal is to enhance users' communication performance in a fair manner. To achieve this, we address the max-min fairness problem, aiming to maximize the minimum SINR among all users. This objective is achieved by jointly optimizing the hybrid beamforming and phase shift design and power allocation. This approach ensures that each user receives a satisfactory level of SINR, promoting fairness in the system while efficiently utilizing the available resources. The optimization problem can be formulated as:

$$\max_{c_u, c_e, \mathbf{W}, \mathbf{\Theta}} \quad \gamma_i^{t,min}, \tag{29}$$

subject to:

$$\begin{aligned}
&\mathbf{S_1} : \gamma_k^{m,t} > \gamma_{th}, \\
&\mathbf{S_2} : \sum_{k=1}^{K} c_k = 1, \\
&\mathbf{S_3} : \beta_n^t + \beta_n^r = 1; \;\; \beta_n^t, \beta_n^r \in [0,1], \\
&\mathbf{S_4} : \theta_n^t, \theta_n^r \in [0, 2\pi], \\
&\mathbf{S_5} : \; \|\mathbf{w}_m\|_F^2 = 1, \quad \forall m, \\
&\mathbf{S_6} : \; |A| = \frac{1}{\sqrt{N}}.
\end{aligned}$$

Constraint $\mathbf{S_1}$ demonstrates that the SINR of eMBB and uRLLC users must be higher than threshold values (γ_{th}) to satisfy the minimum QoS requirements. The sum of transmission power gains of NOMA clusters must be one according to constraint $\mathbf{S_2}$. Constraints $\mathbf{S_3}$ and $\mathbf{S_4}$ represent the constraints of transmission and reflection of the STAR-RIS. Finally, constraints $\mathbf{S_5}$ and $\mathbf{S_6}$ are the restriction of the power for the hybrid beamforming matrix and the restriction of constant modulus in the analog domain, respectively.

B. Proposed Solution

In this context, we propose a solution for determining the beamforming matrix at the BS, the phase shift matrix at the STAR-RIS, and power allocation, as outlined in (29). Due to the non-convex nature of the objective functions, conventional convex optimization based approaches become ineffective. Thus, we advocate the use of DRL algorithms, with the following parameters:

Agent: The BS acts as the agent in the DRL framework.

State: The state vector is composed of the latest received SINR feedback of the users, which is expressed as

$$\mathbf{s}_t = [\gamma_u(t), \gamma_e(t)]. \tag{30}$$

Since the optimization objective aims at maximizing the minimum user SINR, these values provide direct information about the instantaneous quality of service experienced by each user. To improve numerical stability during training, the SINR values are normalized as

$$\gamma_{\text{norm}} = \frac{\gamma - \gamma_{\min}}{\gamma_{\max} - \gamma_{\min}}, \tag{31}$$

where $\gamma_{\max}$ and $\gamma_{\min}$ denote the maximum and minimum SINR values observed during training, respectively. This compact state design reduces the dimensionality of the learning space while maintaining effective performance feedback for the control decision process.

Action: The continuous action vector, denoted as $\mathbf{a}_t$, corresponds to the set of available beamforming, STAR-RIS phase shift matrices, and transmission powers, which is defined as

$$\mathbf{a}_t = [c_u, c_e, \mathbf{W}, \mathbf{\Theta}]. \tag{32}$$

where the power allocation coefficients satisfy $0 \leq c_u, c_e \leq 1$, $c_u + c_e = 1$, the beamforming matrix satisfies $\|\mathbf{W}\|_F^2 \leq 1$, and the STAR-RIS phase shifts are constrained as $\beta_n^t, \beta_n^r \in [0,1]$, $\beta_n^t + \beta_n^r = 1$ and $\theta_n^t, \theta_n^r \in (0, 2\pi]$.

Reward: The direction of the agent's movement is dictated by the reward function, formulated based on the objective of the DRL method. Considering (29), the aim of the optimization problem is to maximize the minimum SINR among all users while satisfying the QoS for each slice [47]. Therefore, the reward function ($\mathbf{r}_t$) can be formulated as follows:

$$\mathbf{r}_t = \gamma_i^{t,min} - |\gamma_k^{m,t} - \gamma_{th}| - |\sum_{k=1}^{K} c_k - 1| - |\beta_n^t + \beta_n^r - 1| - |\,\|\mathbf{w}_m\|_F^2 - 1| - |\,|A| - \frac{1}{\sqrt{N}}|. \tag{33}$$

The following presents the details of the proposed solution.

C. Softmax Deep Double Deterministic Policy Gradients (SD3) Algorithm

The SD3 algorithm is a model-free reinforcement learning framework designed for continuous control problems. It follows an actor–critic paradigm, where the objective is to learn an optimal deterministic policy $\pi(s)$ that maximizes the expected long-term discounted reward. In SD3, the policy and action-value functions are approximated by deep neural networks, referred to as the actor and critic, respectively [48].

Let $s_t \in \mathcal{S}$ and $a_t \in \mathcal{A}$ denote the system state and continuous action at time slot t. The critic network estimates the action-value function

$$\mathcal{Q}^{\pi}(s_t, a_t) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} \,\middle|\, s_t, a_t\right], \tag{34}$$

where r_t is the immediate reward and $\gamma \in (0,1)$ is the discount factor. The actor network outputs a deterministic action $a_t = \pi_\theta(s_t)$, parameterized by θ, which aims to maximize the critic's evaluation.

To mitigate overestimation bias commonly observed in standard DDPG, SD3 employs a double critic architecture consisting of two independent $\mathcal{Q}$-functions, $\mathcal{Q}_1(s,a)$ and $\mathcal{Q}_2(s,a)$. The target $\mathcal{Q}$-value is computed using a clipped double $\mathcal{Q}$-learning strategy combined with a Boltzmann softmax operator. Specifically, the target value for critic update is given by

$$y_t = r_t + \gamma\, \mathcal{T}_{\text{soft}}\left(\min_{i=1,2} \mathcal{Q}_i'(s_{t+1}, a_{t+1}')\right), \tag{35}$$

where $\mathcal{Q}_i'(\cdot)$ denotes the target critic networks and $a_{t+1}' = \pi_{\theta'}(s_{t+1})$ is the action generated by the target actor network.

The Boltzmann softmax operator $\mathcal{T}_{\text{soft}}(\cdot)$ is defined as

$$\mathcal{T}_{\text{soft}}(\mathcal{Q}) = \sum_j \frac{\exp(\beta \mathcal{Q}_j)}{\sum_k \exp(\beta \mathcal{Q}_k)}\, \mathcal{Q}_j, \tag{36}$$

where β is the temperature parameter controlling the smoothness of the softmax approximation. This operator provides a smooth estimate of the maximum $\mathcal{Q}$-value, improving stability and robustness against noisy value estimates.

The critic networks are trained by minimizing the mean squared temporal-difference (TD) error

$$\mathcal{L}(\phi_i) = \mathbb{E}_{(s_t,a_t,r_t,s_{t+1})\sim\mathcal{D}}\left[\left(\mathcal{Q}_i(s_t, a_t; \phi_i) - y_t\right)^2\right], \tag{37}$$

where ϕ_i represents the parameters of the i-th critic network and $\mathcal{D}$ denotes the replay buffer.

The actor network is updated using the deterministic policy gradient, expressed as

$$\nabla_\theta J(\theta) = \mathbb{E}_{s_t\sim\mathcal{D}}\left[\nabla_a \mathcal{Q}_1(s_t, a)\big|_{a=\pi_\theta(s_t)} \nabla_\theta \pi_\theta(s_t)\right], \tag{38}$$

which encourages the actor to select actions that maximize the estimated $\mathcal{Q}$-value.

To stabilize training, SD3 maintains target networks for both actor and critic, which are updated using a soft update mechanism

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta', \quad \phi_i' \leftarrow \tau\phi_i + (1-\tau)\phi_i', \tag{39}$$

where $\tau \ll 1$ controls the update rate.

Furthermore, SD3 employs an experience replay mechanism, where past transitions (s_t, a_t, r_t, s_{t+1}) are stored and randomly sampled to decorrelate training data and improve sample efficiency. The complete actor–critic interaction and training procedure are illustrated in Fig. (2), while the step-by-step implementation of the SD3 algorithm is summarized in Algorithm (1).

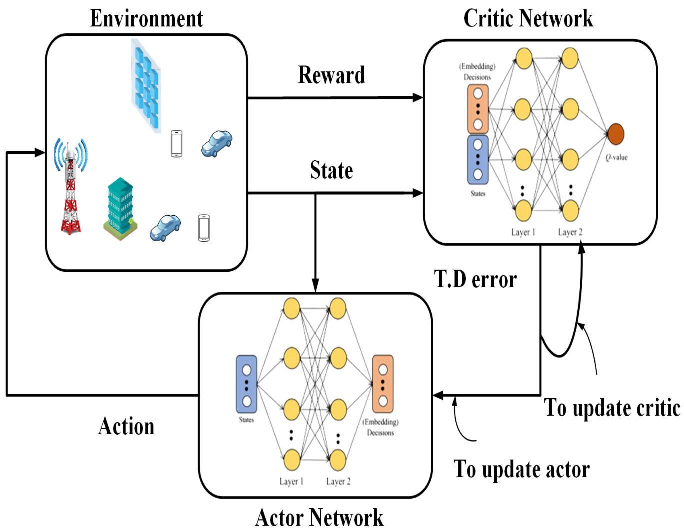


Fig. 2: Block diagram of actor-critic reinforcement learning

Algorithm 1 Pseudo-code of the SD3 Algorithm

Initialize two critic networks $\mathcal{Q}_1(s, a; \theta_1)$, $\mathcal{Q}_2(s, a; \theta_2)$
Initialize two actor networks $\pi_1(s; \phi_1)$, $\pi_2(s; \phi_2)$
Initialize target networks: $\theta_i' \leftarrow \theta_i$, $\phi_i' \leftarrow \phi_i$, $\forall i \in \{1, 2\}$
Initialize replay buffer $\mathcal{D}$
for $t = 1, \ldots, T$ **do**
  Observe current state s_t
  Select action: $a_t = \pi_k(s_t; \phi_k) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma)$ where $k \in \{1, 2\}$ is chosen uniformly
  Execute a_t, observe reward r_t, next state s_{t+1}, and done flag d_t
  Store transition $(s_t, a_t, r_t, s_{t+1}, d_t)$ in $\mathcal{D}$
  Sample a mini-batch of N transitions from $\mathcal{D}$
  for $i = 1, 2$ **do**
    Sample K target action noises $\epsilon_k \sim \mathcal{N}(0, \bar{\sigma})$
    Generate perturbed target actions: $a_{t+1}^{(k)} = \pi_i'(s_{t+1}; \phi_i') + \text{clip}(\epsilon_k, -c, c)$
    Compute clipped double Q-values: $\hat{\mathcal{Q}}^{(k)} = \min_{j=1,2} \mathcal{Q}_j'(s_{t+1}, a_{t+1}^{(k)}; \theta_j')$
    Compute softmax Bellman target: $\mathcal{T}_{\text{soft}} = \frac{\sum_{k=1}^{K} \exp(\beta \hat{\mathcal{Q}}^{(k)}) \hat{\mathcal{Q}}^{(k)}}{\sum_{k=1}^{K} \exp(\beta \hat{\mathcal{Q}}^{(k)})}$
    Compute TD target: $y_t = r_t + \gamma(1 - d_t)\mathcal{T}_{\text{soft}}$
    Update critic by minimizing Bellman loss: $\mathcal{L}(\theta_i) = \frac{1}{N} \sum \left(\mathcal{Q}_i(s_t, a_t; \theta_i) - y_t\right)^2$
    Update actor using deterministic policy gradient: $\nabla_{\phi_i} J = \frac{1}{N} \sum \nabla_a \mathcal{Q}_i(s, a; \theta_i)\big|_{a=\pi_i(s)} \nabla_{\phi_i} \pi_i(s; \phi_i)$
  end for
  Update target networks (soft update): $\theta_i' \leftarrow \tau\theta_i + (1 - \tau)\theta_i', \quad \phi_i' \leftarrow \tau\phi_i + (1 - \tau)\phi_i'$
end for

D. Complexity Analysis

The computational complexity of the proposed DRL-based optimization framework depends on the dimensions of the state and action spaces, as well as the complexity of the learning algorithm. For the considered STAR-RIS-assisted system, the computational complexity of continuous-control DRL algorithms, including DDPG, TD3, and SD3, is approximately expressed as $O(2K^2N^2M)$, where K, N, and M denote the number of users, STAR-RIS elements, and BS antennas, respectively. For comparison, an exhaustive search-based optimization approach requires searching all possible combinations of power allocation, beamforming, and STAR-RIS transmission/reflection coefficients. Therefore, its computational complexity is given by $O(2KN^2ML_PL_WL_\beta L_\theta)$, where L_P, L_W, L_β, and L_θ represent the quantization levels of the transmit power, BS beamforming coefficients, STAR-RIS amplitude coefficients, and phase shifts, respectively. Due to the exponential growth of the search space, exhaustive search becomes impractical for large-scale STAR-RIS deployments.

The computational complexity orders of different optimization approaches are summarized in Table I. Although SD3 introduces additional computational operations compared with DDPG and TD3 due to the utilization of multiple critics and softmax-based critic aggregation, it still provides a significantly lower complexity compared with exhaustive search while achieving improved optimization performance.

TABLE I: COMPLEXITY ANALYSIS

Algorithm	Complexity order
DDPG	2.1×10^6
TD3	3.2×10^6
SD3	3.9×10^6
Exhaustive Search	2.3×10^{10}

Furthermore, the scalability of the proposed framework is affected by the dimension of the continuous action space. The STAR-RIS phase-shift coefficients introduce an action dimension proportional to the number of controllable RIS elements. In addition, the beamforming variables scale with the number of BS RF chains and served users. Therefore, the action dimension approximately grows as $O(N + MN_{\mathrm{RF}})$, where N_{RF} denotes the number of RF chains. Increasing the number of STAR-RIS elements provides additional degrees of freedom for improving system performance; however, it also enlarges the exploration space and increases the training complexity of DRL algorithms.

To further evaluate the practical computational cost, the empirical training time, number of environment interactions, and performance variance are measured under identical simulation settings. Each DRL algorithm is trained using multiple independent random seeds, and the average results are reported in Table II. Although SD3 requires higher training time than DDPG and TD3 due to its additional critic networks, it achieves more stable convergence and higher optimization performance, providing a favorable trade-off between computational cost and system performance.

TABLE II: EMPIRICAL TRAINING COMPLEXITY OF DRL ALGORITHMS

Algorithm	Training Time (min)	Environment Interactions	Std. Dev.
DDPG	42.5	8×10^5	1.8
TD3	58.7	8×10^5	2.4
SD3	71.3	8×10^5	2.1

Finally, practical deployment of STAR-RIS-assisted systems introduces several implementation challenges. In particular, large-scale STAR-RIS architectures require efficient control signaling mechanisms between the BS and the RIS controller to update the transmission and reflection coefficients. Moreover, accurate CSI acquisition for the cascaded BS–STAR-RIS–user channels remains challenging due to the large number of RIS elements and associated training overhead. Therefore, practical implementations require low-overhead channel estimation techniques and efficient RIS control protocols to enable real-time optimization.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed algorithms for beamforming design at the BS, phase shift matrix design at the STAR-RIS, and power allocation. The values of the parameters used in the simulation are listed in Table III.

A. Performance of Different DRL Methods

Here, we delve into the analysis of the sum-rate performance across different scenarios as illustrated in Figs. 3a, 3b, and 3c. These figures provide insights into how the sum-rate of clusters—representing the total data transmission rate across all users within each cluster—varies depending on the DRL methods used, the presence of different system impairments, and the impact of varying transmit power. In Fig. 3a, we explore the sum-rate performance under the assumption of perfect SIC, where higher-priority signals are perfectly decoded, and interference is completely removed for subsequent signals. We compare the performance of three DRL algorithms: Deep Deterministic Policy Gradient (DDPG), Twin Delayed DDPG (TD3), and SD3. The results clearly show that the SD3 algorithm outperforms both DDPG and TD3 across different levels of transmit power. This superior performance is attributed to SD3's ability to more effectively optimize the beamforming at the BS, the phase shifts of the STAR-RIS, and the transmission power, leading to a maximized sum-rate. In Fig. 3b, we introduce the challenge of imperfect SIC, where the cancellation of interference from other users' signals is not flawless. Here, the imperfect SIC coefficient is set to 0.2 ($\xi = 0.2$), which means that a portion of the interference remains even after the SIC process. Despite this complication, the SD3 algorithm again demonstrates its robustness by achieving a higher sum-rate compared to TD3 and DDPG. The degradation in performance due to imperfect SIC is noticeable, yet SD3 maintains a lead by effectively managing the residual interference and optimizing the system's resources. Fig. 3c addresses the scenario where the system is affected by IQI impairment, a common hardware imperfection that leads to phase and amplitude mismatches. IQI can severely degrade the system's performance by introducing signal distortions. Despite these challenges, SD3 continues to deliver superior sum-rate

TABLE III: SIMULATION PARAMETERS

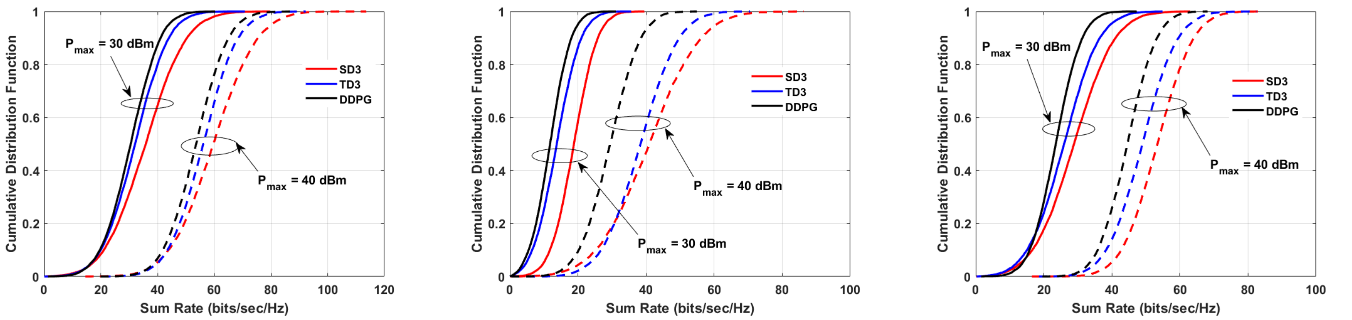
Parameter	Value
Carrier frequency	28 GHz
Bandwidth	100 MHz
Time-slot duration	1 ms
Mini-slot duration	0.125 ms
Max transmission power	30, 40 dBm
Number of BS transmit antennas	16
Number of RF chains	3
Number of data streams	3
Number of RBs	9
Number of clusters	3
Number of UEs	6
Number of STAR-RIS elements	4, 16, 32
Noise power	-174 dBm/Hz
Minimum SINR	2
DRL Training Hyperparameters	
Actor network architecture	FC(512)-FC(256)-FC(128)-FC(action_dim)
Critic network architecture	FC(512)-FC(256)-FC(128)-FC(1)
Activation function	ReLU (hidden layers), tanh (output layer)
Replay buffer size	1×10^6 transitions
Mini-batch size	256
Discount factor (γ)	0.99
Target update rate (τ)	0.005
Optimizer	Adam
Actor learning rate	1×10^{-4}
Critic learning rate	3×10^{-4}
DDPG exploration noise	$\mathcal{N}(0, 0.2^2)$
TD3 policy noise	0.2
TD3 noise clipping	$[-0.5, 0.5]$
TD3 delayed policy update	Every 2 critic updates
SD3 exploration strategy	Softmax-weighted double critics with adaptive noise
SD3 noise decay	$\sigma_t = 0.3e^{-t/50000}$
Warm-up steps	10,000
Training episodes	4000
Steps per episode	200
Total environment interactions	8×10^5
Gradient clipping	Maximum norm = 1

performance compared to the other algorithms. The results indicate that SD3 is better equipped to mitigate the adverse effects of IQI, ensuring that the system can still achieve a high level of data transmission efficiency.

Therefore, across all scenarios—whether with perfect SIC, imperfect SIC, or IQI—SD3 consistently outperforms TD3 and DDPG. This suggests that SD3 is more effective in handling complex optimization problems involving the joint tuning of beamforming, phase shifts, and power allocation. Also, increasing the maximum transmit power positively impacts the SINR of users, leading to an improved sum-rate. This correlation highlights the importance of power control in maximizing network performance, particularly in systems with sophisticated resource allocation strategies like those enabled by SD3.

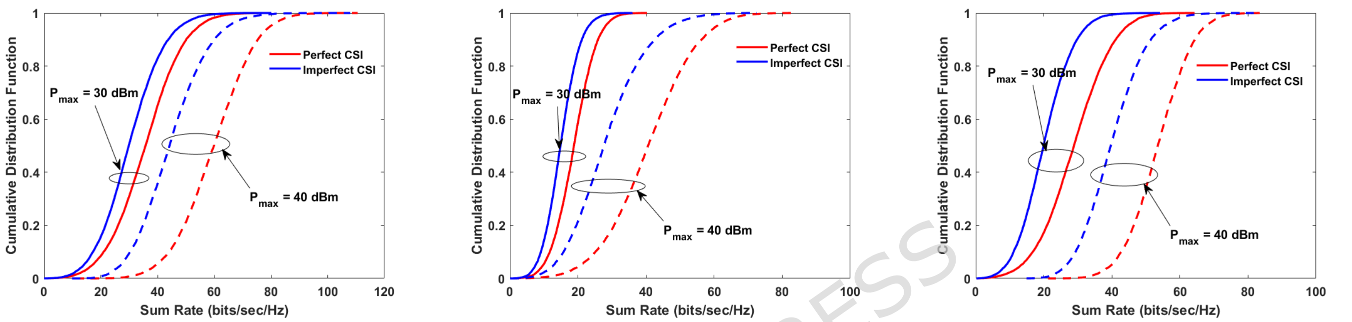
B. Impact of CSI uncertainty

Figs. 4a, 4b, and 4c illustrate the impact of imperfect CSI on the achievable sum rate under three different scenarios: the ideal system without hardware impairments, the system with imperfect SIC, and the system with IQI. In each case, the performance is evaluated for different transmit power levels and CSI accuracy conditions. It can be observed that imperfect CSI causes a noticeable degradation in the sum-rate performance compared with the perfect CSI case. This degradation is caused by the mismatch between the actual channel state and the estimated channel information used for optimizing the STAR-RIS coefficients and beamforming vectors. As the CSI uncertainty increases, the probability of selecting suboptimal transmission parameters increases, resulting in a reduction of the achievable sum rate. Furthermore, the impact of CSI errors becomes more pronounced in the presence of additional system impairments. In the imperfect SIC scenario, channel uncertainty further increases the residual interference after SIC decoding, which leads to a larger performance loss. Similarly, under IQI, imperfect CSI aggravates the distortion caused by the image signal component, resulting in additional degradation of the achievable sum rate.



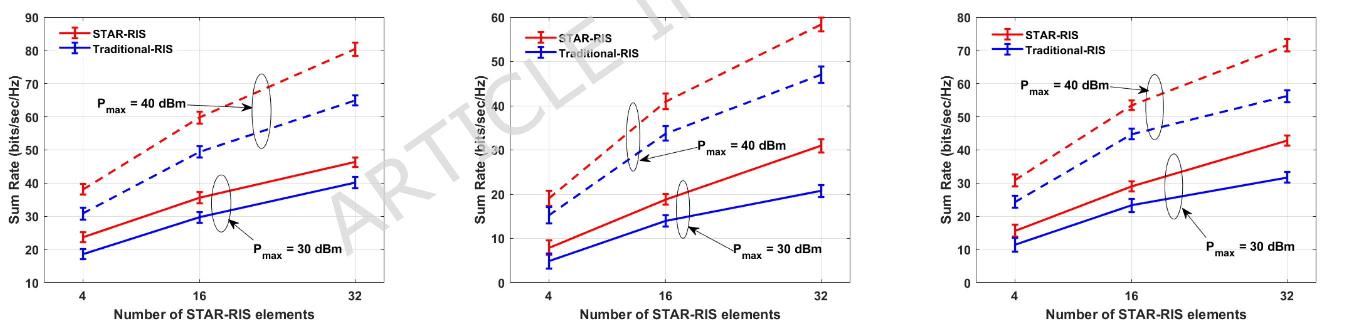
(a) (b) (c)

Fig. 3: Performance comparison of different DRL methods for $N_{SR} = 16$, $\xi = 0.2$, $g_t = g_r = 0.8$, and $\phi_t = \phi_r = 5°$, a) Perfect system, b) Imperfect system, c) IQI impairment



(a) (b) (c)

Fig. 4: Impact of CSI uncertainty for $N_{SR} = 16$, $\xi = 0.2$, $g_t = g_r = 0.8$, and $\phi_t = \phi_r = 5°$, a) Perfect system, b) Imperfect system, c) IQI impairment

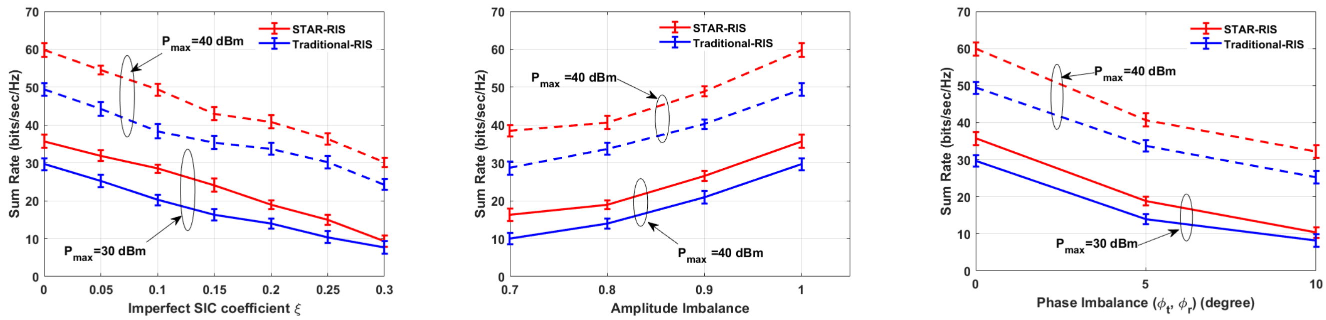


(a) (b) (c)

Fig. 5: The achievable sum-rate versus the number of STAR-RIS elements with $\xi = 0.2$, $g_t = g_r = 0.8$, $\phi_t = \phi_r = 5°$, and SD3 algorithm, a) Perfect system, b) Imperfect system, c) IQI impairment

C. Impact of the Number of STAR-RIS Elements

Figs. 5a, 5b, and 5c, illustrate the performance of our proposed STAR-RIS system compared to a traditional RIS-assisted setup utilizing SD3 algorithm. The key metric under consideration is the achievable sum-rate, which represents the total data rate at which all users whithin a network. The figures show how this sum-rate varies with the number of RIS elements (N_{SR}) under different conditions: perfect SIC, imperfect SIC, and IQI impairments. In Fig. 5a, we assume that SIC is performed perfectly, meaning that the interference from stronger users' signals is completely removed for subsequent users. The figure shows that as the number of RIS elements (N_{SR}) increases, the achievable sum-rate improves for both the STAR-RIS and traditional RIS systems. This improvement is expected because more RIS elements enhance the ability to manipulate and direct the signal, leading to better signal strength and coverage. However, the STAR-RIS system consistently outperforms the traditional RIS setup. The conventional system uses separate RISs for transmission and reflection, which limits its ability to optimize signal paths fully. In contrast, the STAR-RIS system can dynamically adjust both transmission and reflection at the same location, leading to a more effective utilization of available resources and a higher sum-rate. In Fig. 5b, we introduce the challenge of imperfect SIC, where interference cancellation is not flawless, leaving some residual interference



(a) (b) (c)

Fig. 6: Sensitivity analysis of hardware impairments: a) imperfect SIC coefficient, b) IQI amplitude mismatch, and c) IQI phase mismatch.

in the system. Despite this complication, the overall trend remains similar: an increase in the number of RIS elements (N_{SR}) leads to a higher sum-rate for both systems. However, the performance gap between STAR-RIS and traditional RIS systems becomes more pronounced under imperfect SIC. This is because the STAR-RIS system's ability to simultaneously manage both transmission and reflection allows it to better mitigate the effects of residual interference, thereby preserving a higher sum-rate compared to the traditional setup, which is constrained by its separate RISs. In Fig. 5c, the scenario considers the presence of IQI impairment, a common hardware imperfection that causes mismatches in phase and amplitude. IQI typically degrades system performance by introducing signal distortions. Even under these challenging conditions, increasing N_{SR} still leads to an improved sum-rate for both systems, but again, the STAR-RIS system significantly outperforms the traditional RIS setup. This superior performance is due to the STAR-RIS's enhanced flexibility in managing signal distortions caused by IQI, optimizing the signal paths more effectively than a traditional RIS system, which has limited control over the transmission and reflection process. Therefore, Fig. 5 highlights the benefits of STAR-RIS over traditional RIS systems, showcasing its ability to achieve higher sum-rates through better resource optimization and increased robustness against common system impairments.

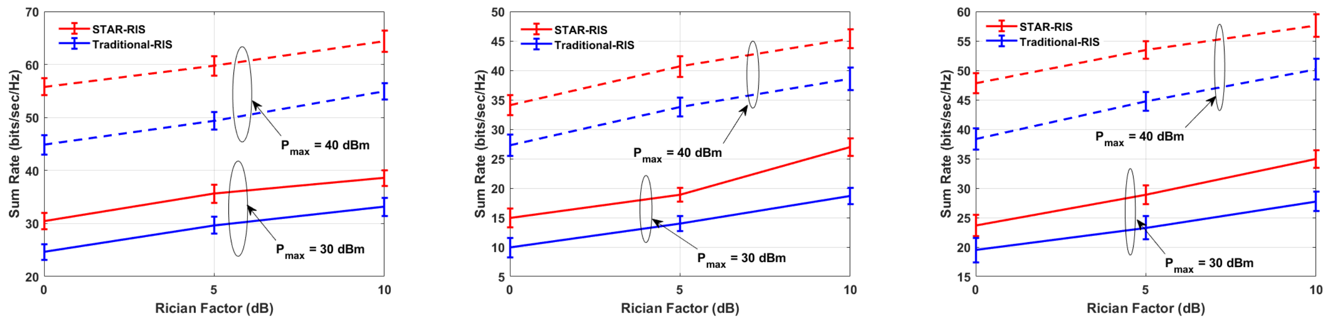
D. Impact of Hardware Impairments

To further evaluate the robustness of the proposed STAR-RIS-assisted framework, this subsection investigates the impact of practical hardware impairments on the system performance. Specifically, three representative impairment parameters are analyzed independently: the imperfect SIC coefficient, the IQI amplitude mismatch, and the IQI phase mismatch. In all experiments, the proposed SD3-based optimization is considered, and the performance of the STAR-RIS is compared with that of a conventional RIS. Fig. 6a illustrates the impact of the imperfect SIC coefficient ξ, where the IQI parameters are fixed at $g_t = g_r = 0.8$ and $\phi_t = \phi_r = 5^\circ$. A larger value of ξ indicates stronger residual interference after SIC. As expected, the achievable sum rate decreases as ξ increases because the residual interference degrades the decoding performance of the NOMA users. Nevertheless, the proposed STAR-RIS consistently outperforms the conventional RIS over the entire range of SIC impairment levels. Fig. 6b presents the effect of the IQI amplitude mismatch by varying the transmitter and receiver amplitude mismatch factors while fixing the phase mismatch at $\phi_t = \phi_r = 5^\circ$. As the amplitude mismatch becomes more severe, the image interference introduced by IQI increases, leading to SINR degradation and a corresponding reduction in the achievable sum rate. Despite this performance degradation, the proposed STAR-RIS maintains a clear advantage over the conventional RIS under all considered amplitude mismatch levels. Finally, Fig. 6c shows the influence of the IQI phase mismatch by varying the transmitter and receiver phase imbalance while fixing the amplitude mismatch at $g_t = g_r = 0.8$. Increasing the phase mismatch strengthens the mirror-image interference generated by the I/Q branches, resulting in additional signal distortion and reduced system throughput. Although the sum rate decreases with increasing phase imbalance, the proposed STAR-RIS consistently achieves higher performance than the conventional RIS, demonstrating its improved robustness to IQI impairments.

Overall, these results confirm that both imperfect SIC and IQI degrade the achievable sum rate. However, the proposed STAR-RIS-assisted SD3 framework exhibits greater resilience to practical hardware impairments and consistently provides superior performance compared with the conventional RIS.

E. Impact of Channel Propagation Conditions

Fig. 7 illustrates the impact of the Rician factor on the achievable sum rate, where the SD3 algorithm is employed for resource optimization. As the Rician factor increases, the dominant LoS component becomes stronger, while the contribution of the scattered NLoS component decreases. Consequently, the effective BS STAR-RIS user channel becomes more reliable, resulting in higher received SINR and improved sum-rate performance for both the proposed STAR-RIS and the conventional RIS schemes. Furthermore, the proposed STAR-RIS consistently outperforms the conventional RIS across the entire range of Rician factors, demonstrating its ability to more effectively exploit favorable propagation conditions through simultaneous transmission and reflection.



(a) (b) (c)

Fig. 7: Impact of the Rician factor: a) Perfect system, b) Imperfect system, c) IQI impairment.

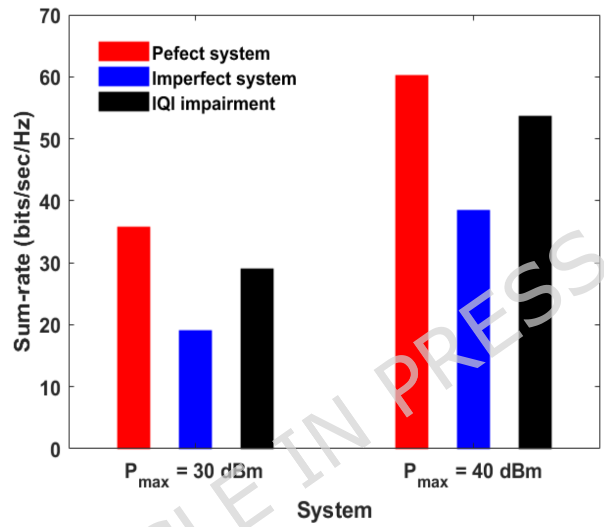


Fig. 8: Performance Comparison of different scenarios with $N_{SR} = 16$, $\xi = 0.2$, $g_t = g_r = 0.8$, $\phi_t = \phi_r = 5°$ and SD3 algorithm

F. Performance Comparison of Different Effects

In the analysis of different scenarios affecting system performance, it is evident from Fig. 8 that perfect SIC outperforms both IQI and imperfect SIC. Perfect SIC represents an ideal scenario where all interference from stronger signals is completely removed before decoding weaker signals. This scenario allows the system to operate under optimal conditions, resulting in the highest achievable sum-rate or data transmission efficiency. The absence of residual interference in perfect SIC ensures that each user receives their intended signal without any significant distortion, leading to superior system performance.

On the other hand, when comparing IQI and imperfect SIC, the system performance under IQI conditions proves to be better than that under imperfect SIC. Although IQI introduces hardware-induced distortions due to phase and amplitude imbalances, these distortions have a less severe impact on the system compared to the residual interference left by imperfect SIC. The interference caused by imperfect SIC can be more unpredictable and damaging, leading to a greater reduction in overall system performance. Therefore, while it is important to address hardware imperfections like IQI, the findings suggest that improving SIC processes to minimize residual interference should be a key focus in system design to achieve closer-to-optimal performance levels.

G. Convergence Analysis

The convergence analysis of DRL algorithms for the proposed problem is illustrated in Fig. 9. Among the evaluated algorithms, DDPG exhibited the fastest convergence, stabilizing at an optimal solution more quickly than both TD3 and SD3. Convergence, in this context, refers to the algorithm's ability to reach a stable or optimal policy during training. A faster convergence means that the algorithm identifies this optimal solution more quickly. While DDPG converged faster, SD3 showed slower convergence but ultimately demonstrated better performance than both DDPG and TD3. Despite taking longer to stabilize, SD3's final results surpass the others, indicating a trade-off between speed and long-term performance. This suggests that while DDPG might be preferable when quick training is crucial, SD3 offers superior performance with extended training.

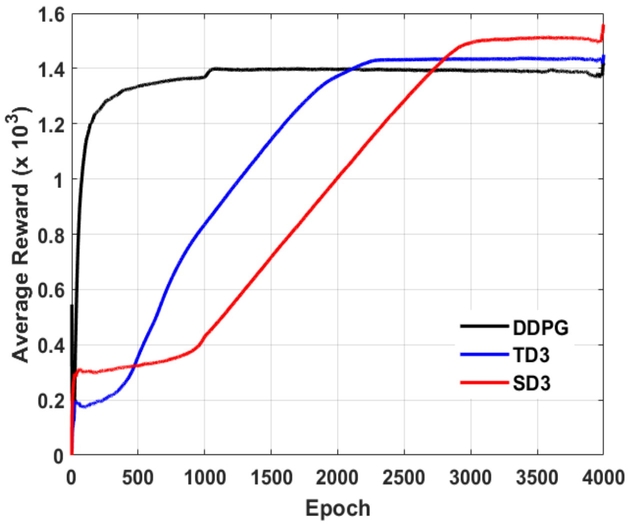


Fig. 9: Convergence of DRL methods.

V. Conclusion

In this paper, we investigated the enhancement of hybrid mmWave-NOMA networks through the integration of STAR-RIS and DRL. A joint optimization framework was developed to optimize power allocation, hybrid beamforming, and STAR-RIS phase-shift design with the objective of maximizing the minimum SINR among users. Simulation results demonstrated that the proposed SD3-based optimization consistently outperformed DDPG and TD3 in terms of convergence behavior and achievable cluster sum rate. Moreover, STAR-RIS achieved significantly higher performance than conventional RIS, particularly at higher transmit power levels, highlighting the benefit of simultaneous transmission and reflection. The robustness of the proposed framework was further evaluated under several practical impairments. The impact of imperfect CSI showed that channel estimation errors degrade the achievable sum rate due to the mismatch between the estimated and actual channels. In addition, a comprehensive sensitivity analysis was conducted for hardware impairments, including the imperfect SIC coefficient, IQI amplitude mismatch, and IQI phase mismatch. The results demonstrated that increasing the residual SIC interference or the severity of IQ imbalance reduces the system performance; however, the proposed SD3-based STAR-RIS framework consistently maintained superior performance and exhibited greater robustness than the benchmark schemes. Furthermore, the impact of the Rician factor was investigated, showing that stronger LoS propagation improves the achievable sum rate for both STAR-RIS and conventional RIS, while STAR-RIS consistently preserves a clear performance advantage across different propagation environments.

Overall, the obtained results demonstrate that the proposed STAR-RIS-assisted SD3 optimization framework provides an effective and robust solution for practical mmWave-NOMA systems operating under realistic channel conditions and hardware impairments, offering valuable insights for the design of future beyond-5G and 6G wireless networks.

CRediT authorship contribution statement

Sima Sobhi-Givi: Writing – original draft, Methodology, Formal analysis, Conceptualization, Software, Investigation.
Mahdi Nouri: Writing – review & editing, Validation, Formal analysis, Supervision.
Mousa Abdollahvand: Conceptualization, Methodology, Writing – review & editing.
Hamid Behroozi: Validation, Writing – review & editing.
Zhiguo Ding: Supervision, Writing – review & editing, Conceptualization.
Zahriladha Zakaria: Validation, Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sima Sobhi-Givi reports a relationship with the University of Mohaghegh Ardabili, Department of Electrical Engineering, that includes academic affiliation. The authors declare that there are no conflicts of interest related to this submission. The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

The authors received no specific funding for this work.

REFERENCES

[1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6g wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[2] W. Shi, W. Xu, X. You, C. Zhao, and K. Wei, "Intelligent reflection enabling technologies for integrated and green internet-of-everything beyond 5g: Communication, sensing, and security," *IEEE Wireless Communications*, vol. 30, no. 2, pp. 147–154, 2023.

[3] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[4] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE journal on selected areas in communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[5] M. D. Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable ai meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.

[6] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6367–6383, 2022.

[7] H. Ozpinar, S. Aksimsek, and N. T. Tokan, "A novel compact, broadband, high ggain millimeter-wave antenna for 5g beam steering applications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 2389–2397, March 2020.

[8] W. Wu, D. Liu, X. Hou, and M. Liu, "Low-complexity beam training for 5g millimeter-wave massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 361–376, January 2019.

[9] F. Dong, W. Wang, and Z. Wei, "Low-complexity hybrid precoding for multi-user mmwave systems with low-resolution phase shifters," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 9774–9784, October 2019.

[10] M. Rihan and L. Huang, "Optimum co-design of spectrum sharing between mimo radar and mimo communication systems: An interference alignment approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 667–11 680, December 2018.

[11] S. Blandino, C. Desset, A. Bourdoux, and S. Pollin, "Energy efficiency of multiple-input, multiple-output architectures: Future 60-ghz applications," *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 65–71, June 2020.

[12] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, "Rate-splitting unifying sdma, oma, noma, and multicasting in miso broadcast channel: A simple two-user rate analysis," *IEEE Wireless Communications Letters*, vol. 9, no. 3, pp. 349–353, March 2019.

[13] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X.-G. Xia, "Millimeter-wave noma with user grouping, power allocation and hybrid beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5065–5079, November 2019.

[14] I. Budhiraja, S. Tyagi, S. Tanwar, N. Kumar, and M. Guizani, "Cross layer noma interference mitigation for femtocell users in 5g environment," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4721–4733, May 2019.

[15] S. Sobhi-Givi, M. G. Shayesteh, and H. Kalbkhani, "Energy-efficient power allocation and user selection for mmwave-noma transmission in m2m communications underlaying cellular heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9866–9881, 2020.

[16] S. Sobhi-Givi, M. Nouri, M. G. Shayesteh, H. Kalbkhani, and Z. Ding, "Reinforcement learning based joint resource allocation and user fairness optimization in mmwave-noma hetnets," in *2023 31st International Conference on Electrical Engineering (ICEE)*, 2023, pp. 781–786.

[17] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network ris-assisted network slicing resource optimization algorithm for coexistence of embb and urllc slicing based 5g and future mobile networks: mobility, resource management, and challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, Augest 2017.

[18] I. Chih-Lin, S. Han, Z. Xu, S. Wang, Q. Sun, and Y. Chen, "New paradigm of 5g wireless internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 474–482, March 2016.

[19] B. Matthiesen, O. Aydin, and E. A. Jorswieck, "Throughput and energy-efficient network slicing," in *22nd International ITG Workshop on Smart Antennas (WSA)*, Bochum, Germany, March 2018.

[20] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Mobile network architecture evolution toward 5g," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 84–91, May 2016.

[21] M. Nouri, S. Sobhi-Givi, H. Behroozi, M. G. Shayesteh, M. J. Piran, and Z. Ding, "Joint slice resource allocation and hybrid beamforming with deep reinforcement learning for noma-based vehicular 6g communications," *IEEE Transactions on Network and Service Management*, vol. 22, no. 4, pp. 3162–3178, 2025.

[22] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5g: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, May 2017.

[23] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, October 2018.

[24] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, October 2017.

[25] W. Shi, J. Yao, W. Xu, J. Xu, X. You, Y. C. Eldar, and C. Zhao, "Combating interference for over-the-air federated learning: A statistical approach via ris," *IEEE Transactions on Signal Processing*, vol. 73, pp. 936–953, 2025.

[26] J. Zuo, Y. Liu, Z. Ding, and L. Song, "Simultaneously transmitting and reflecting (star) ris assisted noma systems," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.

[27] V. D. P. Souto, S. Montejo-Sánchez, J. L. Rebelatto, R. D. Souza, and B. F. Uchôa-Filho, "Irs-aided physical layer network slicing for urllc and embb," *IEEE Access*, vol. 9, pp. 163 086–163 098, 2021.

[28] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "Star-riss: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Communications Letters*, vol. 25, no. 9, pp. 3134–3138, 2021.

[29] C. Wu, Y. Liu, X. Mu, X. Gu, and O. A. Dobre, "Coverage characterization of star-ris networks: Noma and oma," *IEEE Communications Letters*, vol. 25, no. 9, pp. 3036–3040, 2021.

[30] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (star) ris aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3083–3098, 2022.

[31] M. Soleymani, I. Santamaria, E. A. Jorswieck, and B. Clerckx, "Optimization of rate-splitting multiple access in beyond diagonal ris-assisted urllc systems," *IEEE Transactions on Wireless Communications*, vol. 23, no. 5, pp. 5063–5078, 2024.

[32] M. Soleymani, I. Santamaria, E. Jorswieck, and S. Rezvani, "Noma-based improper signaling for multicell miso ris-assisted broadcast channels," *IEEE Transactions on Signal Processing*, vol. 71, pp. 963–978, 2023.

[33] M. Soleymani, I. Santamaria, and E. A. Jorswieck, "Rate splitting in mimo ris-assisted systems with hardware impairments and improper signaling," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 4580–4597, 2023.

[34] X. Shen, Z. Zeng, and X. Liu, "Ris-assisted network slicing resource optimization algorithm for coexistence of embb and urllc," *Electronics*, vol. 11, no. 16, p. 2575, 2022.

[35] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.
[36] A. Ranjha and G. Kaddoum, "Urllc facilitated by mobile uav relay and ris: A joint design of passive beamforming, blocklength, and uav positioning," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4618–4627, 2021.
[37] D. C. Melgarejo, C. Kalalas, A. S. de Sena, P. H. Nardelli, and G. Fraidenraich, "Reconfigurable intelligent surface-aided grant-free access for uplink urllc," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*. IEEE, 2020, pp. 1–5.
[38] W. R. Ghanem, V. Jamali, and R. Schober, "Joint beamforming and phase shift optimization for multicell irs-aided ofdma-urllc systems," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1–7.
[39] M. Almekhlafi, M. A. Arfaoui, M. Elhattab, C. Assi, and A. Ghrayeb, "Joint resource allocation and phase shift optimization for ris-aided embb/urllc traffic multiplexing," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1304–1319, 2022.
[40] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: A multi-sector mode enabling highly directional full-space wireless coverage," *IEEE Journal on Selected Areas in Communications*, 2023.
[41] M. Nouri, H. Behroozi, H. Bastami, M. Moradikia, A. Jafarieh, A. Abdelhadi, and Z. Han, "Hybrid precoding based on active learning for mmwave massive mimo communication systems," *IEEE Transactions on Communications*, vol. 71, no. 5, pp. 3043–3058, 2023.
[42] J. Zuo, Y. Liu, Z. Ding, and L. Song, "Simultaneously transmitting and reflecting (star) ris assisted noma systems," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.
[43] E. N. Tominaga, H. Alves, R. D. Souza, J. L. Rebelatto, and M. Latva-Aho, "Non-orthogonal multiple access and network slicing: Scalable coexistence of embb and urllc," in *IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. Helsinki, Finland: IEEE, April 2021.
[44] S. Sobhi-Givi, M. Nouri, M. G. Shayesteh, H. Behroozi, H. H. Kwon, and M. J. Piran, "Efficient optimization in ris-assisted uav system using deep reinforcement learning for mmwave-noma 6g communications," *IEEE Internet of Things Journal*, vol. 12, no. 14, pp. 26 042–26 057, 2025.
[45] W. Shi, J. Xu, W. Xu, C. Yuen, A. Lee Swindlehurst, and C. Zhao, "On secrecy performance of ris-assisted miso systems over rician channels with spatially random eavesdroppers," *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 8357–8371, 2024.
[46] X. Tian, Q. Li, X. Li, H. Peng, C. Zhang, K. M. Rabie, and R. Kharel, "I/q imbalance and imperfect sic on two-way relay noma systems," *Electronics*, vol. 9, no. 2, p. 249, 2020.
[47] S. Sobhi-Givi, M. Nouri, M. G. Shayesteh, H. Kalbkhani, and Z. Ding, "Joint power allocation and user fairness optimization for reinforcement learning over mmwave-noma heterogeneous networks," *IEEE Transactions on Vehicular Technology*, pp. 1–16, 2024.
[48] L. Pan, Q. Cai, and L. Huang, "Softmax deep double deterministic policy gradients," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11 767–11 777, 2020.