IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 3, Ser. III (May. – June. 2026), PP 01-12
www.iosrjournals.org

Deep Reinforcement Learning for RIS-Assisted NOMA Systems for 6G Networks

Rupinder Kaur[1], Dr. Mandeep Kaur[2], Dr. R.S. Uppal[3]
[1](M. Tech Student, Department of Electronics & Communication Engineering, BBSBEC, Fatehgarh Sahib, India)
[2](Assistant Professor, Department of Electronics & Communication Engineering, BBSBEC, Fatehgarh Sahib, India)
[3](Professor, Department of Electronics & Communication Engineering, BBSBEC, Fatehgarh Sahib, India)

Abstract:
In this paper, we introduce an approach based on Deep Reinforcement Learning to optimize the RIS-NOMA scheme for the 6G wireless network. Some other parameters considered as actions in the proposed algorithm are clustering of users, power distribution, and RIS phase optimization. The two types of deep reinforcement learning algorithms, namely, SAC and MADDPG, were implemented on the proposed architecture. The performance of the proposed framework has been examined based on some parameters including sum rate, SINR, throughput, delay, energy efficiency, and Jain's fairness index. The suggested MADDPG and SAC models have shown better performance than existing models, based on the obtained results.

Key Word*: Deep Reinforcement Learning (DRL), Non-Orthogonal Multiple Access (NOMA), Soft Actor Critic (SAC), Reconfigurable Intelligent Surface (RIS).*

Date of Submission: 02-06-2026 Date of Acceptance: 13-06-2026

I. Introduction

Evolution of 6G Wireless Networks and RIS Technology: The objective behind the implementation of 6G wireless networks is that it will provide extremely fast data transmission speeds, high capacity, and worldwide reach. 6G does not regard the wireless environment as passive but rather considers it an active and programmable environment. The Reconfigurable Intelligent Surfaces contribute immensely by taking advantage of metamaterials in managing how signals propagate with the help of software-defined phase shifts [7], [25]. With RIS, it becomes possible to compensate for double path loss and increase connectivity in low-signal areas [19]. Thus, 6G wireless technology requires smart
Radio environments [7],[19],[25].

Synergizing RIS with NOMA: For achieving higher spectral efficiency, RIS is employed along with NOMA. NOMA supports concurrent access by several users for the same time and frequency resources through varying power allocation and SIC [4], [13]. With the integration of RIS, interference can be managed and utilized effectively, thus enhancing performance in dense scenarios [28]. With respect to this approach, both spectral efficiency and SIC can become more efficient. Such integration is important for supporting many users with good QoS, especially at the network edge [4],[13],[28].

Challenges in Resource Allocation and Move to DRL: But there is another issue with this scheme—resource management will become quite challenging. The optimal resource management is not a linear optimization problem and is thus hard to implement using traditional methods such as fractional programming and the Lagrange method [16], [26]. In order to cope with real-time requirements in 6G systems, DRL-based solutions are adopted. Techniques like SAC and MADDPG are capable of adjusting for channel variability and Rayleigh Fading [7],[14],[26].

Research Contribution: This study attempts to enhance the management of resources in 6G communication networks through artificial intelligence. The main contributions are the following:

- **Joint Optimization:** This is where the researchers develop a system to enhance the power control of base stations and phase changes of RIS simultaneously for improved performance of the system.
- **Usage of Advanced AI Approaches:** Two advanced approaches to SAC, and MADDPG, have been used to solve problems of the complex 6G networks.
- **Multi-Agent Collaboration:** This has enhanced collaboration between the agents in the system to deliver enhanced performance in the system.

- **Outperforms Existing Approaches:** The proposed technique has been found to outperform the existing technique in terms of data rate, delay, and quality of signals.
- **Balanced Performance:** In addition to enhancing the efficiency of the network, the developed system has managed fairness in performance.

II. Literature Review

Classical Optimization vs. Heuristic Methods: In the initial studies, the resource allocation problem in the RIS-enabled NOMA communications was addressed through mathematical optimization approaches. SCAs and block coordinate descent were two approaches employed in phase and power allocation [17], [18]. Though they give the optimal solution, they are highly complex and depend highly on CSI, which is hard to acquire in 6G communication environments [1], [18]. Heuristic approaches came up as another alternative, but they face issues of slow convergence rate and tend to be trapped in local minima [1], [17].

Deep Reinforcement Learning (DRL) in Wireless Networks: The advent of the DRL technique offers another way to solve optimization problems involving a non-convex resource allocation problem. Early work considered learning from the perspective of single-agent reinforcement learning (RL), specifically through Deep Q-Networks (DQN) for resource block distribution in RIS-aided NOMA systems [21], [22]. Also, the Policy Gradient (PG) method can be effective in dealing with continuous action space problems, e.g., RIS phase optimization [10]. Although there has been considerable progress in the field of DRL, various challenges still remain to be addressed, particularly with respect to exploration/exploitation balance in dense user clusters [10],[21],[22].

Comparative Analysis and Research Gap:

Author & Year	Method Used	Core Technique	Optimization Goal	Remarks
Zhong et al. (2020) [26]	DRL	Deep Q-Network (DQN)	Sum Rate Maximization	Limited to discrete phases
Hou et al. (2020) [4]	Optimization	SCA & Successive Interference Cancellation	Outage Probability	High computational complexity
Yang et al. (2021) [20]	DRL	DDPG	Energy Efficiency	Unstable in high mobility
Mu et al. (2021) [15]	Traditional Opt.	Block Coordinate Descent	Spectral Efficiency	Requires impractical perfect CSI
Wang et al. (2021) [16]	Machine Learning	Federated Learning	QoS & Latency	High communication overhead
Zhang et al. (2022) [24]	DRL	Proximal Policy Optimization (PPO)	Secure Throughput	Struggled with multi-user interference
Li et al. (2023) [9]	Hybrid DRL	Double DQN + DDPG	Fair Power Allocation	Lacks multi-agent coordination
Chen et al. (2024) [2]	Multi-Agent DRL	Independent Q-Learning	Network Capacity	Poor global fairness
Proposed Work (2026)	**Joint DRL**	**SAC & MADDPG**	**Throughput, Energy Efficiency, Fairness, Latency**	**Uses centralized critics to ensure cooperative behaviour and robust 6G adaptability.**

Table 1: Comparing Literature Review and Proposed Research Contribution

Current researches have either concentrated on the development of single-agent DRL or the optimization of some parameters only. Multi-agent collaboration is still relatively unexplored.

III. System Model & Mathematical Formulation

Network Model: The design of the network is that of a single antenna base station (BS), which transmits to four users (K=$User_1, User_2, User_3, User_4$) in power-domain NOMA downlink communication. The users are divided into 2 clusters (M) to manage interference using NOMA. In this setup, the direct signal from the BS to the users is prevented. Hence, RIS can be employed to reflect and enhance the reflected signals via providing another communication channel. RIS consists of N = 16 passive reflecting elements. All element n ϵ {1, 2, ……, 16} has the ability to change its phase. The response of the RIS is represented by the diagonal matrix $\boldsymbol{\Phi} = \text{diag}\left(e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}\right)$, where $\theta_N \epsilon [0, 2\pi]$ denotes the phase change of the n-th element [5],[24].

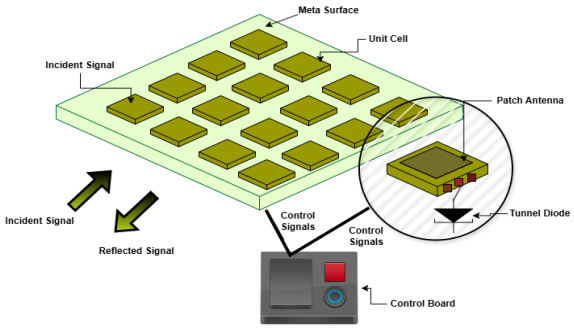


Figure 1: Schematic representation of a passive RIS element.

Table 2: Simulation Parameter

Parameter	**Value**
Total Users (K)	4
Total RIS Elements (N)	16
NOMA Clusters (C)	2
Small-Scale Fading	Rayleigh Fading
Action Dimension	20
State Dimension	12
Proposed Algorithm	SAC, MADDPG
Number of Training Episodes	600
Steps per Episodes	30

Channel Modelling: The communication environment incorporates three distinct propagation links:
1. Line of Sight (Direct Path) (BS → User): Represented by $h_{d,k} \in \mathbb{C}^{1*1}$.
2. Reflection via RIS Path 1 (BS → RIS): Represented by $h_{d,r} \in \mathbb{C}^{N*1}$.
3. Reflection via RIS Path 2 (RIS → User): Represented by $h_{r,k} \in \mathbb{C}^{1*N}$.

In order to emulate the randomness of 6G, Rayleigh fading is used to simulate small-scale fading and it describes multipath scenario where there is no dominant LoS component [9],[24]. The effective channel gain for user k is formulated as:

$$h_{total} = h_{direct} + h_{RIS} \tag{1}$$

The received signal y_k at user k can be written as:

$$y_k = h_{total}s + n \tag{2}$$

h_{direct} stands for the path from base station to the user directly, while h_{RIS} is the cascaded channel through RIS (base station to RIS to user). s is the transmitted signal, and n is the noise added to the received signal. The double-path loss constraint is observed where the attenuation of the signal control is based on the product of the distances of two-hop reflection channels [5],[23].

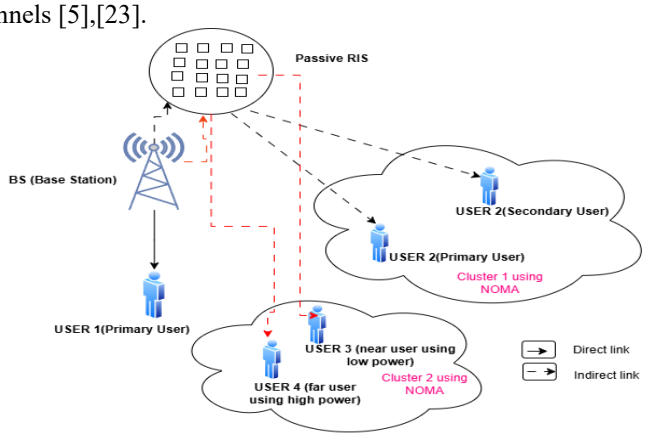


Figure 2: Framework Architecture for a 6G Wireless System using the concept of Passive RIS to facilitate communication for users with NOMA.

NOMA Transmission and SIC Assumptions: Power-domain NOMA is used in which the BS broadcasts the superposition signal x=$\sum_{k=1}^{K}\sqrt{p_k}\,s_k$, where p_k is the power assigned to user k. For the sake of realism, the following assumptions are adopted:

- **Perfect SIC:** It is considered that the Successive Interference Cancellation (SIC) process is executed without error by all receivers in accordance with the order of effective channel gains $|h_{eff,1}|>|h_{eff,2}|$ [3], [23].
- **Passive Reconfigurable Intelligent Surface (RIS):** The individual components of RIS will be considered as passive elements, thus changing the phase of signals without increasing their signal power [5], [9].
- **Quasi-Static Channels:** The state of the channel is supposed to remain static within one episode step but vary randomly between the steps [2],[11].

In figure 3, User 2 (far user) is allotted a higher power level (P_1) while User 2 (near user) is allotted a lower power level (P_2) on the same frequency subcarrier. The SIC decoding procedure for the receiver side is illustrated in the following figure on the right-hand side.

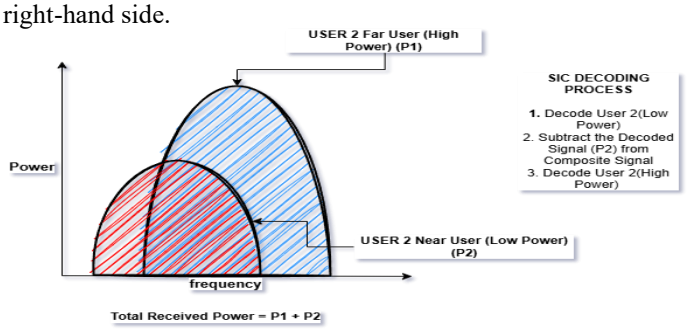


Figure 3: Principle of NOMA & SIC

Network Parameters: The following system of measurement are used to evaluate system performance:

- **Signal to Interference plus Noise Ratio (SINR):** For NOMA-based system, SINR of user k is calculated through Successive Interference Cancellation (SIC). If it assume that users have been sorted on the basis of the magnitude of effective channel ($|h_{eff,1}|^2 \geq |h_{eff,2}|^2$ ....). The SINR (γ_k) is expressed as:

$$\gamma_k = \frac{P_k|h_{total}|^2}{\sum_{j>k}P_j|h_{total}|^2 + \sigma^2} \tag{3}$$

  The above equation, P_k refers to the transmission power and n_o is additive white Gaussian noise power.
- **Sum Rate and Throughput:** The sum rate (R_{Sum}) is the total data rate of all users in the cluster, derived from the Shannon capacity formula. Throughput is the practical data rate (in kbps or Mbps) calculated from the sum rate.

$$R_{Sum} = Bandwidth * log_2(1 + \{SINR\}) \tag{4}$$

- **Network Latency:** As the rate increases, the latency decreases.

$$L = \frac{\eta}{1 + \zeta.R} \tag{5}$$

  L stands for latency (ms), R represents the sum rate (bps/Hz), and $\eta = 1.2$ is the base delay constant. The scaling parameter for latency with respect to the system rate is $\zeta = 0.1$.
- **Energy Efficiency:** Energy efficiency (η_{ee}) is the measure of energy required by the NOMA network in the form of a ratio:

$$\eta_{ee} = \frac{R_{Sum}}{P_{RIS} + hardware\ Power} \tag{6}$$

- **Jain's Fairness Index:** It shows how uniformly resources are allocated to users:

$$\text{Jain's Fairness Index (J)} = \frac{(\sum_{k=1}^{K} R_k)^2}{K.\sum_{k=1}^{K} {R_k}^2} \tag{7}$$

The key ranges from 1/K (worst case) to 1 (ideal fairness), ensuring a balanced quality of service (QoS) across the network.

IV. Problem Formulation

In this paper, an intelligent RIS-based NOMA approach is presented to optimize the performance of the system. This includes optimizing the power distribution, user grouping, and RIS phase shift settings. Optimization in these factors will result in improvements in SINR, sum rate, throughput, energy efficiency, fairness, and reduction of the latency period.

Objective Function: The purpose of the suggested RIS-based NOMA scheme is to optimize the performance of the network by ensuring good signal transmission, decreasing latency, and increasing energy efficiency and fairness.

$$max_{P,\Theta} \sum_{k=1}^{K} (log_2 (1 + SINR_k) - (L_k + E_k + F_k) \quad (8)$$

The mathematical notation $\log_2(1 + \mathrm{SINR}_k)$ the throughput that can be achieved by each user, whereas L_k, E_k, and F_k for delay, energy efficiency, and fairness, respectively. Such an objective allows for optimal resource allocation via user clustering, power distribution, and phase shifting of RISs in the 6G network environment.

V. Proposed Methodology

In this study, a DRL-based approach is presented to tackle the joint optimization problem of RIS-aided NOMA systems. Two advanced algorithms, namely Soft Actor-Critic and Multi-Agent Deep Deterministic Policy Gradient, are used to improve system performance in terms of throughput, fairness, and energy efficiency. Such solutions are suitable in order to solve the complex task within the dynamic wireless communication environment for 6G [3],[5],[9].

DRL Framework Components: The problem of resource allocation is formulated as a Markov Decision Process (MDP) such that the agent learns to perform actions optimally according to the present state of the network for maximizing future rewards [9].

- **State Space (S):** The state space refers to a multi-dimensional feature space that represents important environmental characteristics like channel quality, level of interference, and the status of the RIS system. The state space provides inputs to the learning agents, allowing them to change their action depending on the state of the network [3].

  $$S = \left\{ \left| h_{eff,k} \right|^2, \gamma_k, P_k \right\}, \forall k \quad (9)$$

  Channel Gains: Instantaneous Channel State Information (CSI) of the direct link between BS and user as well as the reflected links via RIS.
  Power Levels: The current power allocation of each user in the NOMA cluster.
  SINR Levels: SINR level of each user used to determine QoS performance.

Here, a 12-dimensional state space was used, which comprised 4 parameters representing the channel gain for the 4 users, 4 parameters representing the SINR for each of the users, 2 parameters relating to noise, and 2 parameters concerning RIS element parameters.

- **Action Space (A):** The action space is described as either continuous or discrete multi-dimensional vectors by means of which an agent can manipulate parameters of the system. The action space usually comprises parameters like transmission power assignment and the phasing setting on the RIS for RIS-assisted NOMA communication systems. The action space values lie between [-1,1].
  Power Allocation: Power distribution among NOMA users from the base station. $[-1,1] \rightarrow [0, P_{\max}]$
  RIS Phase Shifts: Angle of each RIS element that allows for improving the signals. $[-1,1] \rightarrow [0,2\pi]$

  $$A = \{P_1, P_2, P_3, P_4, \theta_1, \theta_2, \dots, \theta_{16}\} \quad (10)$$

  Such that:
  $P_k \in [0, P_{\max}]$ is the power allocation parameters
  $\theta_n \in [0,2\pi]$ is the phase shift parameter

In the present study, the action space is considered to be a 20-dimensional continuous vector, where each dimension corresponds to RIS phase shift values (16 dimensions) and power allocation (4 dimensions). Every action is always normalized to [-1,1] range [3].

- **Reward Function (R):** This numeric value is obtained from the environment following each action taken by the reinforcement learning agent, based on how good or bad an action was. The reward helps the agent learn how to behave optimally based on the maximization of accumulated reward. In order to improve all network parameters evenly, it can define the reward as:

  $$R = R_{Sum} + T + \eta_{ee} + J + \sum_{k=1}^{K} \gamma_k - L \quad (11)$$

  R_{Sum}represents sum rate, T denotes throughput, η_{ee} is energy efficiency, J indicates fairness, γ_k is SINR of the *k-th* user, and L denotes latency.

This reward definition makes sure that the performance is balanced in the network [3].

Soft Actor-Critic (SAC): SAC is an off-policy DRL method which enhances exploration through the maximization of reward and entropy. It works effectively in continuous action spaces such as RIS phase control

[9], [24]. SAC prevents getting stuck in local minima and promotes learning stability in wireless communication scenarios [24].

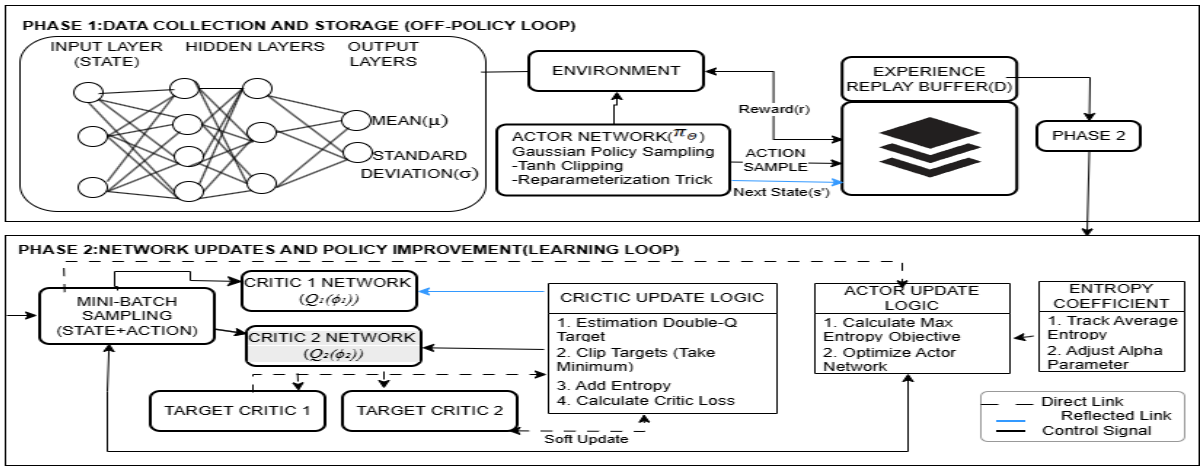


Figure 4: Soft Actor Critic (SAC) DRL Framework architecture flow

The above diagram can be considered split into two parts based on operations performed: Phase 1 (Interaction), which demonstrates the agent interacting with the wireless environment by taking observations from the input layer, processing this information through hidden layers of neurons, and generating action parameters (mean and standard deviation) to allocate resources. Phase 2 (Optimization) shows the process of learning where the Replay Buffer provides data to the Twin-Critics network to determine the Q-values while the Actor Update Logic maximizes the entropy and reward values

Table 3: Hyperparameters of SAC

Hyperparameter	SAC Value
Number of Hidden Layers	2
Activation Function	ReLU (In Hidden Layer) Tanh (in Output Layer)
Number of Neurons in Hidden Layer	[256, 256]
Learning Rate (LR)	3 * 10-4
Optimizer	Adam
Batch Size	64
Discount Factor (γ)	0.99
Soft Update (τ)	0.01
Replay Buffer Size	200,000
Entropy Coefficient Alpha (α)	0.2
Warm-up Step	2,000

Algorithm: 1 (SAC)

Initialization:
Initialize RIS-NOMA environment
Initialize actor network $\pi(\theta)$
Initialize two critic networks $Q_1(\phi_1)$, $Q_2(\phi_2)$
Initialize target networks Q_1', Q_2'
Initialize replay buffer B
Set entropy temperature α and discount factor γ

Training Process:
for each episode do
Reset environment and observe initial state (s)
for each time step do
Sample action $a \sim \pi(a|s)$
Execute action a in the environment
Observe reward r and next state(s')
Store transition (s, a, r, s') in replay buffer B
Sample a mini-batch from replay buffer
Compute target value
Update critic networks Q_1 and Q_2 by minimizing loss
Update actor network π by maximizing expected reward and entropy
Soft update of target networks:
Update state: s = s'
if episode ends then exit loop

end for
end for
Return trained policy and performance metrics

MADDPG (Multi-Agent Framework)

MADDPG is a multi-agent DRL approach based on centralized training and decentralized execution. It allows multiple users to coordinate efficiently in RIS-NOMA systems [3], [8]. MADDPG facilitates collaboration between the users and minimizes interferences, providing higher levels of fairness and system efficiency [3],[8].

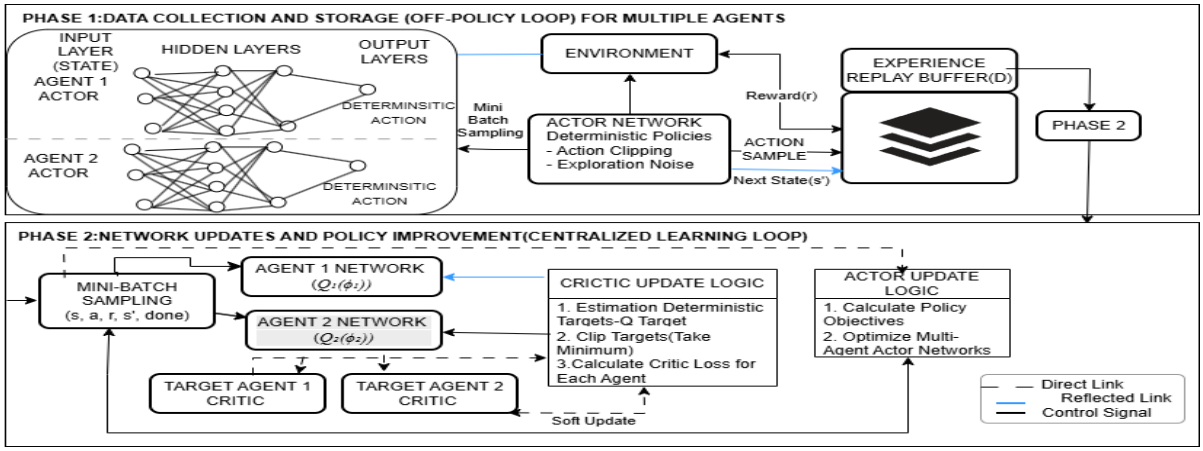


Figure 5: Structural architecture of the MADDPG algorithm, demonstrating a two-stage off-policy paradigm in wireless communications enabled by multiple RISs.

Figure 5 explains Stage 1: decentralized data collection using multi-agent actor networks providing deterministic actions; and Stage 2: centralized learning approach using twin critics to assess joint actions to derive target values for improving actors' policies.

Table 4: Hyperparameters of MADDPG

Hyperparameter	**MADDPG Value**
Number of Hidden Layers	2
Activation Function	ReLU (In Hidden Layer) Tanh (in Output Layer)
Number of Neurons in Hidden Layer	[256, 256]
Learning Rate (LR)	1 * 10-3
Optimizer	Adam
Discount Factor (γ)	0.99
Soft Update (τ)	0.01
Replay Buffer Size	200,000
Noise (Exploration)	0.1

Algorithm 2: MADDPG

Initialize environment env
Initialize actor and critic networks for all agents
Initialize target networks
Initialize replay buffer B
for episode = 1 to 600 do
Reset environment
Observe initial states $s_1, s_2, \ldots, s_n$
for t = 1 to max_steps do
for each agent i do
Select action a_i *using policy* $\pi_i(s_i)$
end for
Execute joint action $(a_1, a_2, \ldots, a_n)$
Observe reward r and next states $s'_1, s'_2, \ldots, s'_n$
Store (s, a, r, s') in replay buffer B
Sample mini-batch from B
Compute target:
$y = r + \gamma * Q\_target(s', a')$

Update centralized critic network
Update each agent's actor network
Soft update target networks
Update states $s_i = s'_i$
if done then
break
end if
end for
end for
Return trained multi-agent policies

Baseline Algorithm: Heuristic Random Resource Allocation

The base algorithm relies on heuristic resource allocation in which the base station (BS) and the reconfigurable intelligent surface (RIS) lack any learning capabilities. This differs from SAC and MADDPG since it is unable to respond to changing channels like Rayleigh fading [8],[29].

Zhou et al. (2023) explained the heuristic methods are widely adopted in RIS-enabled wireless networks because of their simplicity and speed. Nonetheless, they often yield suboptimal performance as opposed to learning techniques [27].

Functional Mechanism:

- **Static RIS Phase Shift:** The phase shifts of the RIS are randomly assigned. This causes ineffective signal combining and failure to reduce double path loss [27].
- **Fixed Power Allocation:** Equal power/fixed ratio of power allocations are adopted without consideration of the Signal-to-Interference-plus-Noise-Ratio (SINR) and Signal Interference Cancellation (SIC). This results in poor allocation of resources [6].
- **Non- learning:** No reward or feedback mechanism is present in the system, hence no improvements with time [12],[27].

Algorithm:3

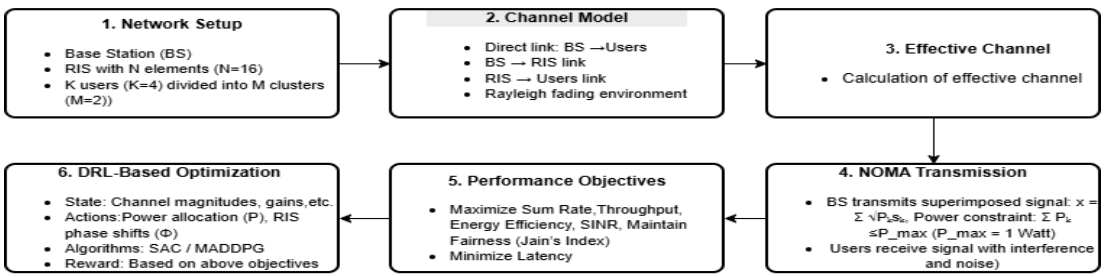
Initialize RIS-NOMA environment
Evaluation Process:
for each episode do
Reset the environment
Initialize performance metrics (e.g., sum rate, energy efficiency, fairness, latency) to zero
for each time step do
Generate a random action a ∈ [-1, 1]
(Action represents random power allocation and RIS phase shifts)
Apply the action to the environment
Observe reward and system performance metrics
Accumulate the metrics over time steps
if episode ends then exit loop
end for
Compute average metrics over all steps in the episode
end for
Return:
Average performance metrics across all episodes

The baseline approach executes non-adaptive and random optimization and operates below optimality. Although straightforward and quick, it leads to increased latency and reduced spectral efficiency than DRL algorithms such as SAC and MADDPG [27].

VI. Simulation Setup

Simulation Tools and Environment: The suggested reinforcement learning schemes (SAC and MADDPG) are implemented in a powerful computing system. The environment itself is designed completely from scratch in order to model the interaction between the RIS and NOMA systems.

FLOW CHART



Programming Language: *Python version 3.10* was used since it is equipped with multiple useful libraries for scientific computing and machine learning.
Hardware Configuration: The simulations were performed on a system with an AMD Ryzen5 7530U processor (2.00 GHz) and 16 GB RAM. The machine is working under the 64-bit Windows 11 Home Operating System. Simulations were done in a CPU-only setting without any GPU acceleration.
Deep Learning and Optimization: *PyTorch (torch, torch.nn*) is used to build and train Actor–Critic networks. *torch.optim* is employed to use the Adam optimization algorithm to update the model's weights.
Environment and Signal Modeling: The *Gymnasium* framework is used to build the customized RIS-NOMA environment with *reset()* and *step()* methods. *NumPy* package is utilized for modelling channels, Rayleigh fading, SINR, and sum rate computations.
Utilities and Visualization: *Matplotlib* library is employed for plotting sum rate and latency against 600 episodes. *Random* and *OS* libraries are used for randomness and file handling.

VII. Results and Discussion

- **Throughput and Sum Rate Analysis:** The throughput and sum-rate performance results reveal a marked enhancement in spectral efficiency.

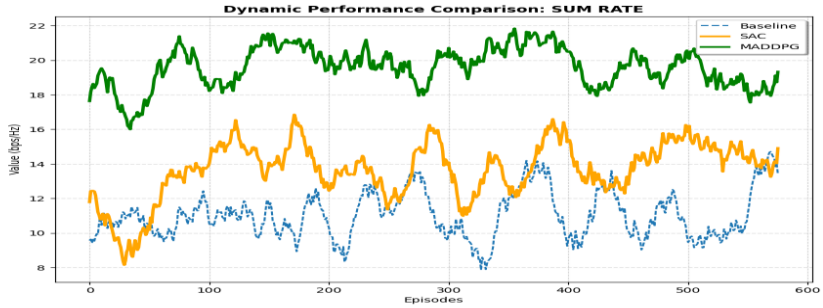


Figure 6: Shows a comparative performance evaluation of the sum rate (bps/Hz) of the network through 600 training episodes.

The above results are derived from running the simulation for 600 episodes. The MADDPG technique obtains an average sum rate of ***20.0 bps/Hz*** as against ***10.5 bps/Hz*** of the baseline and has shown an improvement of ***90.5%.*** Similarly, the SAC technique yields us a mean sum rate of ***13.5 bps/Hz,*** which means a ***28.6%*** improvement.
Through the optimization of 16 RIS elements to reduce interference, the MADDPG algorithm can enhance throughput to ***168 kbps*** from the ***92kbps*** achieved by the baseline system. The SAC algorithm can also improve throughput to ***118 kbps.*** This is obtained through 600 simulation episodes.

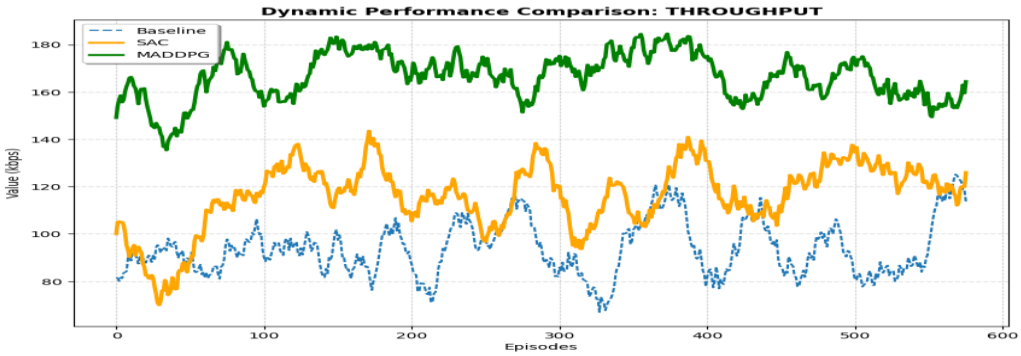


Figure 7: Shows a comparison of system throughput performance (kbps) using various DRL systems.

- **Latency and Energy Efficiency:** Indeed, MADDPG greatly decreases the latency and increases energy efficiency, which are important factors for the performance of 6G systems. The latency drops down to ***0.34ms***, providing a decrease of ***35.8%*** from the base value of ***0.53ms***. Moreover, SAC provides a slight improvement in the reduction of latency to ***0.46ms.***

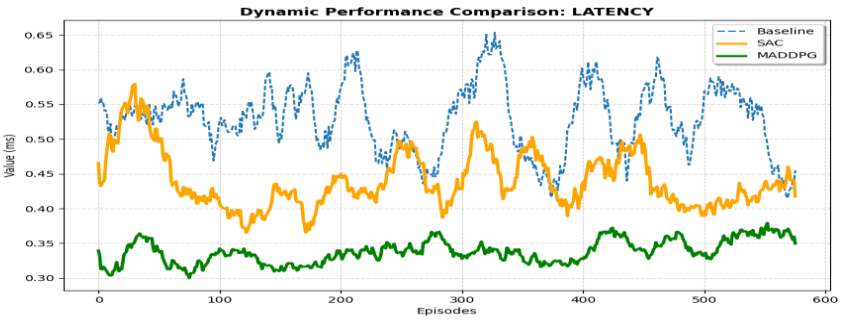


Figure 8: Shows mean network latency (ms) performance, indicating a decrease in transmission time.

In addition, the energy efficiency provided by MADDPG reaches ***5.7 bits/joule***, demonstrating a positive trend in comparison with the base value of ***4.6 bits/joule*** with ***23.91%*** improvement. SAC delivers ***5.4 bits/joule*** of energy efficiency, which is higher than the baseline energy efficiency.

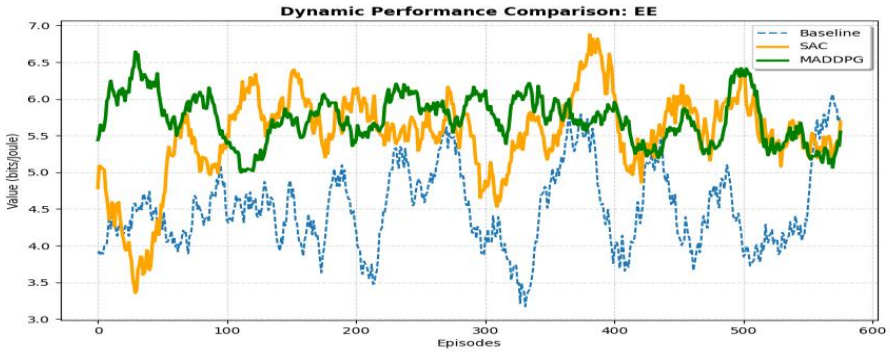


Figure 9: Shows evaluation of energy efficiency (bits/joule) showing the "Green 6G" effect of optimized RIS

- **Signal Quality (SINR) and User Fairness:** The MADDPG algorithm is able to produce high levels of SINR, which averages out to ***11.2 dB*** and represents a gain of ***166.67%*** when compared to the baseline of ***4.2 dB***, allowing efficient use of SIC in NOMA system. The algorithm SAC can achieve a level of SINR of ***7.2 dB***, which although better than the baseline level, still does not match that of MADDPG. As far as fairness is concerned, the benchmark model has an index score of ***0.78*** for Jain, and ***0.80*** for MADDPG, whereas SAC has managed to improve fairness up to ***0.71***.

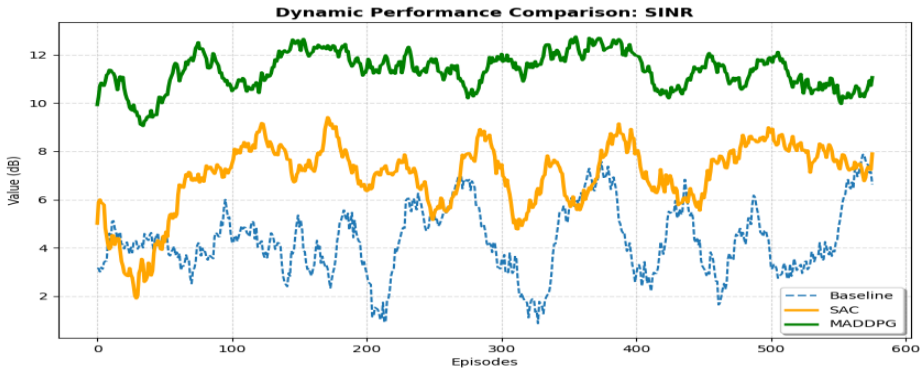


Figure 10: SINR Improvement Comparison with the Assistance of RIS in NOMA Clusters.

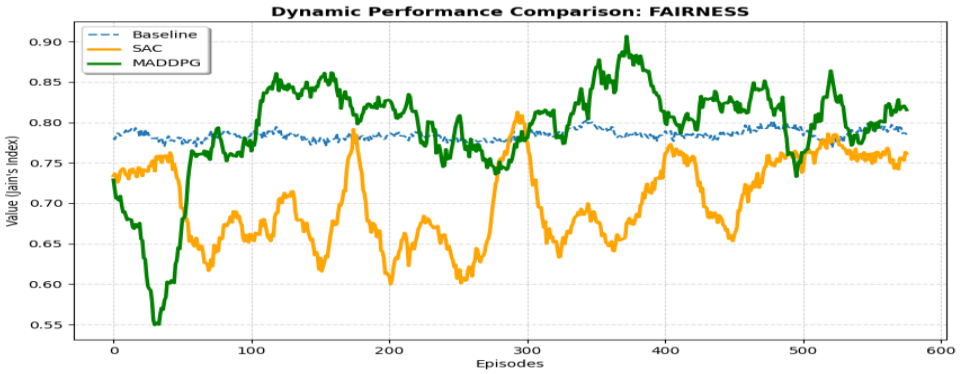


Figure 11: Comparison of Jain's Fairness Index to Evaluate Resource Allocation Equality

Comparative Performance Summary Table

Metric	Baseline (Heuristic) [27]	SAC (Proposed)	MADDPG (Proposed)
Sum-Rate (bps/Hz)	10.5	13.5	20.0
Throughput (kbps)	92.0	118.0	168.0
Latency (ms)	0.53	0.46	0.34
SINR (dB)	4.2	7.2	11.2
Energy Efficiency (EE) (bits/Joule)	4.6	5.4	5.7
Fairness (Jain's Index)	0.78	0.71	0.80

Table 4: Comparison between average values of baseline vs. SAC vs. MADDPG

The proposed intelligent network simulation was done on the specified computer hardware system using the Python programming language. It is found that the parameters yield more efficient results than the traditional optimization techniques.

VIII. Conclusion

The proposed research in this paper aims at creating a system that can intelligently manage resources in NOMA-assisted by RISs for 6G communications. Some of the deep reinforcement learning techniques like SAC and MADDPG will be used to solve challenging problems of power management and RIS phase control while considering important performance metrics.

Future Work

Some improvements can be explored in the future:

- **Mobility Scenarios:** Examine the performance of the system under mobility such as communication between vehicles (V2X communication).
- **Practical Factors:** Take into account practical factors such as partial controllability of the RIS and limitations of hardware.
- **Scalability:** Generalize the model for many users and multiple RIS elements.
- **Security:** Analyze the security provided by using an RIS against hacking.

References

[1] Basar, E., Di Renzo, M., de Rosny, J., Debbah, M., Alouini, M. S., and Zhang, R., "Wireless Communications Through Reconfigurable Intelligent Surfaces," IEEE Access, vol. 7, pp. 116753–116773, 2019.
[2] Chen, J., Zhang, H., and Letaief, K. B., "Multi-Agent Learning for Resource Management in 6G Heterogeneous Networks," IEEE Journal on Selected Areas in Communications, vol. 42, no. 1, pp. 1–15, Jan. 2024.
[3] Gevez. Y, Y. I. Tek, and E. Basar, "Dynamic RIS partitioning in NOMA systems using deep reinforcement learning," Frontiers in Antennas and Propagation, vol. 2, Art. no. 1418412, 2024, doi: 10.3389/fanpr.2024.1418412.
[4] Hou, Z., Sun, Y., Song, M., and Zhang, R., "Reconfigurable Intelligent Surface Assisted Non-Orthogonal Multiple Access," IEEE Transactions on Wireless Communications, vol. 19, no. 11, pp. 6889–6903, Nov. 2020.

[5] Hou, Z., Sun, Y., Song, M., and Zhang, R., "Reconfigurable Intelligent Surface Assisted NOMA Networks: Power Allocation and Phase Shift Design," IEEE Open Journal of the Communications Society, vol. 3, pp. 1–12, 2022.
[6] Hou, Z., Zhang, C., Liu, Y., and Yuan, X., "Reconfigurable Intelligent Surface Assisted NOMA Networks: Power Allocation and Phase Shift Design," IEEE Open Journal of the Communications Society, vol. 3, pp. 123–137, 2022.
[7] Huang, C., Zappone, A., Alexandropoulos, G. C., Debbah, M., and Yuen, C., "Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication," IEEE Transactions on Wireless Communications, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
[8] Li, R. and Xu, L., "Active RIS-Assisted Uplink NOMA with MADDPG for Remote State Estimation in Wireless Sensor Networks," Sensors, vol. 25, no. 4878, 2025.
[9] Li, X., Liu, Y., Ding, Z., and Poor, H. V., "Resource Allocation for RIS-Assisted NOMA Networks via Deep Reinforcement Learning," IEEE Transactions on Communications, vol. 71, no. 5, pp. 1–14, May 2023.
[10] Lillicrap, T. P. et al., "Continuous Control with Deep Reinforcement Learning," arXiv preprint arXiv:1509.02971, 2015.
[11] Liu, R., Mu, X., Liu, Y., and Ding, Z., "Optimization of RIS-NOMA Systems: A Comprehensive Mathematical Survey," IEEE Open Journal of Antennas and Propagation, vol. 6, pp. 1–20, 2025.
[12] Liu, R., Bennis, M., and Poor, H. V., "A Survey of Traditional vs. Learning-Based Optimization in 6G Wireless Networks," IEEE Access, vol. 12, pp. 45678–45702, 2024.
[13] Liu, Y., Qin, Z., Elkashlan, M., Ding, Z., Nallanathan, A., and Hanzo, L., "Non-Orthogonal Multiple Access for 5G and Beyond," Proceedings of the IEEE, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
[14] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., and Mordatch, I., "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," Advances in Neural Information Processing Systems (NeurIPS), pp. 6379–6390, 2017.
[15] Mu, X., Liu, Y., Guo, L., Lin, J., and Al-Dhahir, N., "Joint Design of Transmit Beamforming and Reflection Matrix in RIS-Aided NOMA Systems," IEEE Communications Letters, vol. 25, no. 5, pp. 1608–1612, May 2021.
[16] Wang, J., Song, M., and Sun, Y., "Joint Resource Allocation in RIS-Aided NOMA Systems: A Machine Learning Approach," IEEE Journal on Selected Areas in Communications, vol. 39, no. 7, pp. 2073–2087, Jul. 2021.
[17] Wu, Q. and Zhang, R., "Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming," IEEE Transactions on Wireless Communications, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
[18] Wu, Q. and Zhang, R., "Joint Active and Passive Beamforming Optimization for Intelligent Reflecting Surface Assisted SWIPT," IEEE Wireless Communications Letters, vol. 8, no. 6, pp. 1812–1816, Dec. 2019.
[19] Wu, Q. and Zhang, R., "Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network," IEEE Communications Magazine, vol. 58, no. 1, pp. 106–112, Jan. 2020.
[20] Yang, H., Mu, X., Liu, Y., and Ding, Z., "Deep Reinforcement Learning-Based Resource Allocation for RIS-NOMA Systems," IEEE Transactions on Vehicular Technology, vol. 70, no. 6, pp. 1–13, 2021.
[21] Ye, H., Li, G. Y., and Juang, B. H., "Deep Reinforcement Learning Based Resource Allocation for V2V Communications," IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.
[22] Ye, H., Li, G. Y., and Juang, B. H., "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," IEEE Wireless Communications Letters, vol. 7, no. 1, pp. 114–117, Feb. 2018.
[23] Zhai, X., Wang, J., Sun, Y., and Song, M., "Joint Beamforming and Power Allocation for RIS-Assisted NOMA Networks," IEEE Access, vol. 12, pp. 1–15, 2024.
[24] Zhang, S., Wang, X., and Huang, C., "Robust Beamforming for RIS-Aided Communications: A PPO-Based Reinforcement Learning Approach," IEEE Wireless Communications Letters, vol. 11, no. 3, pp. 1–5, Mar. 2022.
[25] Zhang, Z., Xiao, Y., Ma, Z., Xiao, M., Ding, Z., Lei, X., Karagiannidis, G. K., and Fan, P., "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 28–41, Sep. 2019, doi: 10.1109/MVT.2019.2921208.
[26] Zhong, R., Liu, Y., Mu, X., and Ding, Z., "Deep Reinforcement Learning for RIS-Assisted NOMA Networks," IEEE Transactions on Communications, vol. 68, no. 10, pp. 6406–6420, Oct. 2020.
[27] Zhou, H., Erol-Kantarci, M., Liu, Y., and Poor, H. V., "Heuristic Algorithms for RIS-Assisted Wireless Networks: Exploring Heuristic-Aided Machine Learning," IEEE Wireless Communications, vol. 30, no. 6, pp. 72–79, Dec. 2023.
[28] Zhu, J., Li, Q., Kang, X., and Zhang, R., "Power Allocation and Design of RIS-Assisted NOMA Networks," IEEE Wireless Communications Letters, vol. 9, no. 12, pp. 2140–2144, Dec. 2020.
[29] Zuo, J., Liu, Y., Ding, Z., and Al-Dhahir, N., "Resource Allocation in RIS-Assisted NOMA Systems: A Heuristic Approach," IEEE Wireless Communications Letters, vol. 10, no. 11, pp. 2458–2462, Nov. 2021.