INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2026, VOL. 72, NO. 2, PP. 1-8
Manuscript received January 31, 2026; revised April 2026. doi: 10.24425/ijet.2026.157929

Q-MOHO: Q-Machine Learning Guided Multi-Objective power optimization for RIS-assisted MIMO-NOMA systems

Suprith P G, Marulasiddappa H B, Prashanth G S, and Narayanaswamy G

***Abstract*—In this study, we introduce Q-MOHO (Q-Machine Learning Guided Multi-Objective Hybrid Optimization) for power allocation in MIMO-NOMA communication systems aided by RIS. To optimize throughput, improve equalization, and reduce power variation, the algorithm adjusts power distribution among three users facilitated by 64 RIS components. Dynamically adjusting with SNR, the optimal power distribution spans from 0.2 at elevated SNR levels to 0.8 at reduced SNRs. The Q-MOHO performance of 5.40 bps/Hz at an SNR of 25 dB is substantial when contrasted with equal power allocation (EPA) and difference of convex optimization (DCO). These results confirm that, in real-world wireless conditions, the Q-MOHO algorithm can effectively acquire optimal strategies and significantly enhance system performance.**

***Keywords*—Q: Automated Instruction; Optimization with Various Goals; Distribution of Power; Reconfigurable Intelligent Surface (RIS); MIMO-NOMA; and Optimization of Efficiency**

I. Introduction

THE swift progress of wireless communication technology has turned addressing the rising demand for higher data rates, enhanced accessibility, and improved reliability into a significant challenge. Non-Orthogonal Multiple Access (NOMA) has emerged as an effective method to enhance spectrum efficiency by allowing multiple users to share the same frequency resources via power domain multiplexing.

Reconfigurable Intelligent Surfaces (RIS) offer unprecedented control over the propagation environment by altering reflection characteristics dynamically, while Multiple-Input Multiple-Output (MIMO) systems enhance throughput by leveraging various spatial distributions. Nevertheless, these ground breaking technologies pose complex resource allocation challenges, particularly when both efficiency and equity must be considered concurrently. Traditional approaches that require full channel state information, such as Equal Power Allocation (EPA) or heuristic-driven convex optimization methods like DCO, often fall short in highly dynamic environments.

This document introduces Q-MOHO, a reinforcement learning framework that autonomously creates optimal power distribution algorithms from machine input to address these issues. The approach utilizes Q-Machine learning to optimize multiple objectives—enhancing system throughput while ensuring equitable access for users—without relying on existing models or assumptions. The algorithm's efficiency is evaluated against conventional allocation methods and analyzed under different channel conditions.

The main achievements of this paper are:

1. Creating a multi-objective optimization problem for MIMO-NOMA systems supported by RIS.
2. Development of a structure founded on Q-machine learning that continuously modifies power distribution rules.
3. Thorough computational analysis demonstrating enhanced commuting effectiveness and fairness.

This study offers valuable insights for future 5G and 6G networking technologies, establishing a foundation for highly intelligent, responsive radio networks.

The increasing requirements of contemporary wireless networks for elevated data rates and effective spectrum usage can now be fulfilled by combining Reconfigurable Intelligent Surfaces (RIS) with Multiple-Input Multiple-Output Non-Orthogonal Multiple Access (MIMO-NOMA) systems. RIS enhances signal propagation by smartly reflecting signals toward specific users, thereby improving bandwidth and accessibility. Yet there are many difficulties due to the dynamic nature of wireless channels and the requirement for effective power distribution. Conventional optimization techniques are computationally demanding and frequently need complete Channel State Information (CSI), which limits their use in real-time scenarios.

This research suggests Q-MOHO, a Q-Machine learning-based framework that adaptively learns optimal power allocation techniques without explicit CSI, as a solution to these problems. Q-MOHO formulates the power allocation problem with the following objectives using an optimization approach with multiple objectives assignment:

- Maximize system throughput,
- Enhance user fairness,
- Minimize power variance.

This paper's remaining sections are organized as follows: The relevant work system model is presented in Section 2, the Q-MOHO method, the optimization problem details, and the Q-Machine learning algorithm are formulated in Section 3, the results and discussion are organized in Section 4 and simulation results are discussed. Section 5 summarizes the article.

This work was supported PESITM, Shivamogga, and Karnataka, India.

Suprith P G is with the PESITM, Shivamogga, and Karnataka, India (e-mail: pgsuprith@gmail.com).

Marulasiddappa H B is with the G M University, Davangere, and Karnataka, India (e-mail: marul.bethur@gmail.com).

Prashanth G S is with the JNNCE, Shivamogga, and Karnataka, India (e-mail: gsp341@gmail.com).

Narayanaswamy G is with the PESITM, Shivamogga, and Karnataka, India (e-mail: nswamy@pestrust.edu.in).

II. Related Work

The combination of cutting-edge optimization algorithms and new wireless technologies like Non-Orthogonal Multiple Access (NOMA) and Reconfigurable Intelligent Surfaces (RIS) has garnered a lot of interest in recent years. The demand for effective resource management in changing wireless settings has spurred the investigation of machine learning and reinforcement learning methods to enhance conventional optimization strategies.

Several studies have focused on power management in MIMO-NOMA systems. To enhance spectral efficiency and ensure user fairness, conventional methods such as Equal Power Allocation (EPA) and Convex Optimization (CO) have been thoroughly investigated [1], [2]. Subsequently applied to large-scale networks, these techniques are computationally costly and frequently depend on precise information about channel state (CSI).

In wireless communication, the idea of using Reinforcement learning, also has gained popularity, especially for issues with ambiguous settings or incomplete observations. Q-Machine learning, for instance, has been used for adaptive beam formation [5], energy-effective scheduling of resources [4], and dynamic spectrum access [3]. RL techniques have recently been expanded to multi-agent settings that take user-base station interactions into account [6].

Furthermore, it has been demonstrated that using RIS as a passive beam forming and signal enhancement method can result in significant improvements in coverage, energy efficiency, and interference mitigation [7-8]. In order to increase performance or reduce the likelihood of an outage, some research has concentrated on joint optimization of phase shifts and transmit power [9]. However, few approaches consider the combination of RIS-assisted architectures with NOMA and learning-based power control.

The proposed Q-MOHO framework builds upon these foundations by integrating Q- Machine learning-based multi-objective optimization in a RIS-assisted MIMO-NOMA system. Unlike prior works that optimize for a single metric or require perfect CSI, Q-MOHO balances multiple objectives—throughput, fairness, and power efficiency—while adapting to varying channel conditions. To the greatest extent of our comprehension, this is one of the earliest attempts to implement a structure powered by reinforcement learning that is specifically designed for power allocation in RIS-enabled NOMA systems, helping to achieve reliable and scalable wireless resource management.

This is [10] Multi-Objective Butterfly Optimization Algorithm (MOBOA) to optimize power distribution in NOMA networks. The suggested approach preserves user equity and energy efficiency while optimizing the system total *rate. It illustrates how performance in next-generation wireless* communication systems is improved through bio-inspired optimization. Authors in [11] assesses NOMA's performance in 5G systems that automatically deploy numerous users. The main goals are to enhance frequency efficiency, user connection, and total network throughput. The study underscores how the automated multi-user implementation of 5G networks improves resource efficiency and connectivity dependability. It [12] specifically focuses on reducing service latency and energy usage in 5G NOMA systems through smart compute offloading. A method based on neural networks is employed to enhance the allocation of tasks between users and edge servers. The method decreases processing delays, enhances system performance, and facilitates dependable low-latency communication in future networks.

Citation [13] this research examines the advancement of NOMA's performance and throughput in cognitive radio networks. It examines the effects of spectrum sharing and power allocation on the performance of secondary users. The research asserts that implementing NOMA improves overall network performance, throughput, and spectrum efficiency.

III. Q-MOHO METHOD

The key stages in the Q-MOHO approach are as follows:

Starting Point: This block illustrates all external data that connects to the network. It contains the current RIS (Reconfigurable Intelligent Surface) parameters, channel state information (CSI), and signal-to-noise ratio (SNR). In order to modify the framework's distribution technique in response to shifting network conditions, such as channel degradation or congestion, the input data is crucial.

Second Phase: Transforms continuous parameters, such as SNR and CSI, into different states that the Q-learning system may utilize. The abstraction of information provides it possible to map observations to a small set of states for effective learning by simplifying circumstances for learning with reinforcement.

Phase Three: The framework's fundamental training element. In order to maximize long-term advantages, it investigates or exploits actions (power allocations) and changes the Q-table based on rewards gained [14]. In order to maximize throughput and fairness, the agent develops an ideal strategy for power allocation by striking a balance between investigation (attempting to execute various operations) and exploitation (applying known best operations).

Fourth Phase: The next step assigns transmit power levels to users by applying the chosen action from the Q-Machine learning agent. Depending on the current state and Q-values, power is allocated among users in a way that optimizes the system's operation.

Fifth Phase: Calculates the Signal-to-Interference-plus-Noise Ratio (SINR) with taking into consideration channel's gains and phase shifts of the RIS. SINR is an important indicator of wireless network quality that affects dependability and attainable rate of data [15].

Sixth Phase: Uses system statistics including attainable rate, equality measure, and variation for power allocation to calculate the incentive. By measuring the benefits of an activity and promoting tactics that increase data rate while preserving fairness and lowering variability, the reward function directs the learning process.

Seventh Phase: Maintains and uses the optimal power utilization strategy that the Q-learning agent has discovered for future broadcasts. By consistently improving and employing efficient communication techniques, it guarantees that the system adjusts to changing channel conditions.

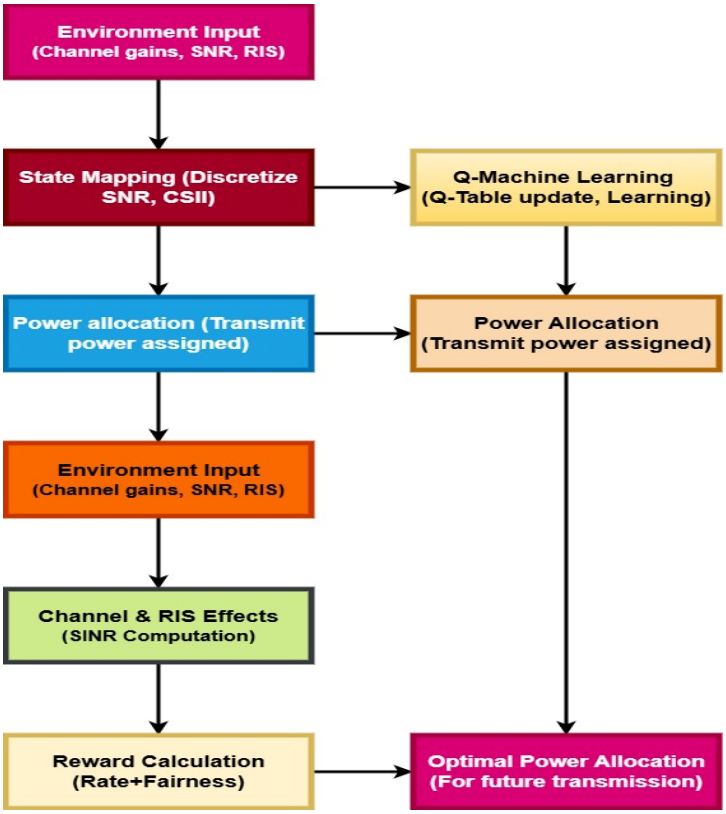


Fig. 1. Block image representing the Q-MOHO Method

3.1 SYSTEM MODEL

In the closely studied downlink MIMO-NOMA systems, a multi-antenna is installed at the base station (BS) to enable service for equipment with multiple users using the same temporal and frequency resources. Consumers are divided into two groups according to how far they are from the BS: proximal and remote. To effectively distribute transmitted power among these two user types, the BS employs a power allocation schema. The signals received by the user terminals are contaminated by "Additive White Gaussian Noise" (AWGN) and suffer from path loss. The system model for the suggested MIMO-NOMA scheme is shown in Fig. 2.

3.2 FORMULATION OF THE PROBLEM

Within the MIMO-NOMA system's wireless communication infrastructure, the planned paradigm is to optimize the power distribution between nearby and remote users. Maximizing the system's achievable aggregate data rate is the main objective. This optimization goal needs to be accomplished while maintaining sufficient amounts of EE and SE at the same time. For the proximal and distant users, let's indicate the allotted Power as u_1 and u_2 such that $u_1 + u_2$ =1. Additionally, the channel gain is defined by the near user as k_1 and k_2 far user. The achievable sum rate s_1 and s_2 are calculated using Shannon capacity theorem [16-17].

$$s_1 = log_2\left[\frac{1+p_t*u_1*k_1}{p_t*u_2*k_1+n_0}\right]$$

$$s_2 = log_2\left[\frac{1+p_t*u_2*k_2}{n_0}\right]$$

Where, p_t is the combined power from BS and log_2 is the logarithm with base 2. The close and distant user's total rate can be calculated by using the expression [18].

$$s_t = s_1 + s_2$$

Accordingly, the optimization task is described as:

Optimization: Maximize the variable-based algorithm u_1 and u_2 provided a set transmission power point while following the guidelines c_1, c_2.

$$s_t = log_2\left[\frac{1+p_t*u_1*k_1}{p_t*u_2*k_1+n_0}\right] + log_2\left[\frac{1+p_t*u_2*k_2}{n_0}\right] \text{ exposed to } c_1, c_2 \quad (1)$$

Where, c_1 demonstrates that all user entities' aggregate transmitted power must equal 1.0. $u_1 + u_2 = 1$. c_2 Ensures that every user in the system has an optimistic power allocation value. The problem outlined in (1) is a constrained optimization challenge, especially when there are more than two customer entities in the system. In these situations, the issue gets increasingly difficult and unsolvable using standard optimization methods. The Q Learning algorithm, a model-free RL methodology, is proposed as a practical solution strategy to address this issue in the current study.

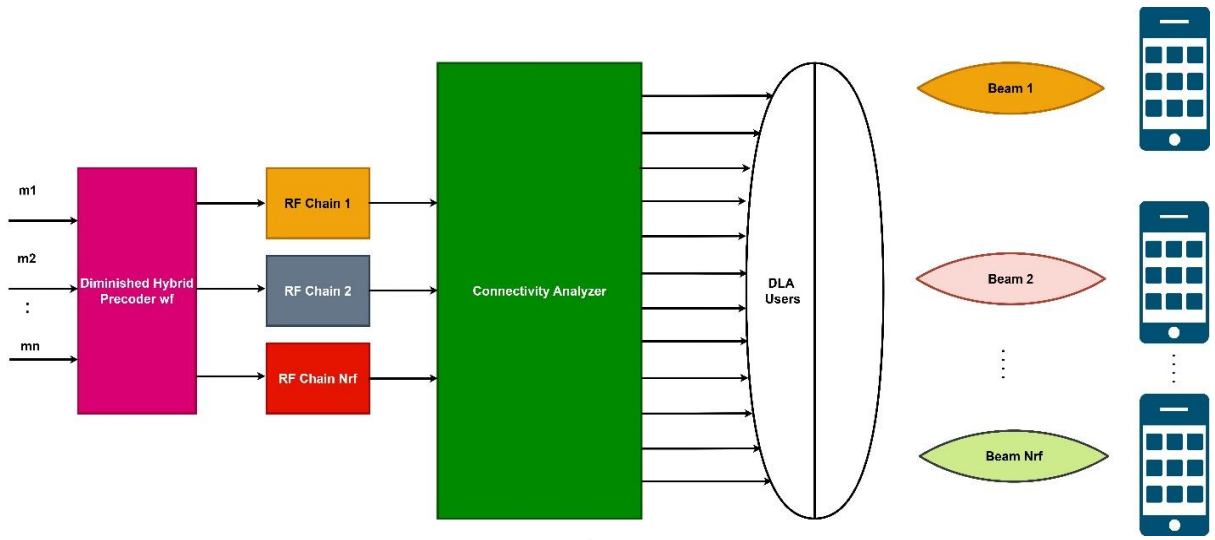


Fig. 2. The suggestion for work's structure model

3.3 Q- MACHINE LEARNING-BASED POWER ALLOCATION ALGORITHM

This section covers Q-learning, a potent model-free technique, and shows how it relates to reinforcement programming (RP), specifically how to extract Q-learning, which is from RP [19-20].

A. The Q- Machine Learning Ideas

An RP model consists of four parameters: the policy, a scalar reward signal, a set of possible environmental conditions, and a set of possible agent behaviors. R, A, and P stand for these four traits.

- Regions of the framework (R)

It explains the conditions of the unmanned aerial vehicle relaying system environment, and decisions about what to do are made in light of the network's current conditions. The number of nodes and their positions, as well as each node's connection and power transmission, are the main variables influencing the network environment. In this approach, the network controller is seen as the learner, receives all of this data so that it may modify the network as necessary to enhance network performance. A countable set is used to define the regions of the framework R as

$$R= R\,(u, e, p) = \{R_0\,, R_1\ \ldots\ldots R_t \ldots\ldots R_T\}$$

We use the € greedy policy for action selection in order to profit from exploration and exploitation [21]. In particular, the operation that optimizes the reaction-value relation in the present state Si is chosen using the likelihood € where € = [0, 1]. In contrast, the action is selected at random from A with a chance of 1 − €. This is done for exploration in order to guarantee the acquisition of the global optimal solution. The following action will be selected at iteration i:

$$R_i = \begin{Bmatrix} arg & \max\limits_{a\in R} & Q\ (R, A)\ with\ probability\ 1-\epsilon \\ & & \mu(R) \qquad with\ probability\ \epsilon \end{Bmatrix}$$

$Q\ (R, A)$ is the state-action function's definition in A_i both the distribution of μ (A) over the set R is uniform.

- Area for activity (A)

Through a series of activities, the learner in the system under consideration attempts to optimize the accumulated incentives, which immediately improves the system's security and outage performance. Reducing the probability of a system hybrid outage is the goal of the optimized work. Consequently, we define the instant reward as the difference between the current and previous system hybrid outages. As the outage decreases, the immediate reward is positive; otherwise, it is negative. Consequently, the instant reward might be offered as [22].

$$A_i = P_{hyt} - P_{hy,(t+1)}$$

Where P_{hyt} is the probability of a hybrid outage at the moment t and $P_{hy,(t+1)}$ is the probability of a hybrid outage at any given time (t +1) following the controller has taken decision R_t to state $R_{(t+1)}$.

Purpose of bonuses (P)

When a student is in state R_t, it has the ability to do something A_t = P (R_t). A learner's goal is to identify the best course of action that maximizes the overall expected reward over the course of operation, which is defined as

$$V^{\pi}(R_t) = \sum_{i=0}^{T_{max}} \beta^{iR}(t+i)$$

Where β € (0, 1) is the reward discount factor. If the value of β is zero, that implies solely instant gratification. R_t is taken into consideration, however, if around 1, it indicates that the reward in the future is more significant than the reward now. The best course of action π^{Δ} is the strategy that is given as follows and maximizes the overall reward [23].

$$\pi^{\Delta} = arg\ max_{A_t} V^{\pi}(R_t), Y(R_t)$$

Substitute the $V^{\pi}(R_t)$ in the above equation.

$$\pi^{\Delta} = arg\ max_{A_t}[R\,(R_t, A_t) + \beta V^{\pi}(R_{t+1}, A_{t+1})]$$

Putting the best possible strategy into action in the equation mentioned above need complete understanding of the state actions data, which is challenging, hence we define [23].

$$P\,(R_t, A_t) = R\,(R_t, A_t) + \beta V^{\pi}(R_{t+1}, A_{t+1})$$

This way

$$\pi^{\Delta} = arg\ max_{A_t} P\ (R_t, A_t)$$

Technique 1: Q Machine Learning-Based Variable Energy Management Technique

Data input: Q Machine Learning α ϵ [0, 1], β ϵ [0, 1], ∈ ϵ [0, 1], Z_1 and Z_2

Data Output: Best course of action π^{Δ}, the best possible degree of power P^{Δ}

1. for t_1= 1 to Z_1 do
2. Decide on a starting state. S_0 at random
3. for t_2= 1 to Z_2 do
4. Set a random number as the initial value μ~ μ [0, 1]
5. if μ>ϵ then
6. use
7. choose an approach of conduct A_t founded on a rapacious approach
8. get rewarded right away R_t and the next state A_{t+1}
9. revise the Q-table to be consistent with (3)
10. network power consumption should be adjusted in accordance with R_t
11. otherwise
12. examine
13. choose a course of action R_t at random
14. get rewarded right away
15. revise the Q-table in accordance with
16. network power distribution should be adjusted in accordance with R_t
17. end if
18. end for
19. end for

Q-Machine learning is typically computed using the iterative technique described below:

$$P(R_t, A_t) = (1-\alpha)\, P(R_t, A_t) + \alpha\, [R(R_t, A_t) + \beta \max(R', A')] \quad (3)$$

B. Q-MACHINE LEARNING-BASED POWER ALLOCATION ALGORITHM

The specifics of the method for allocating power according to Q-Machine Learning is provided Technique 1. During the data entry stage, the networks and Q-Machine learning parameters are initialized. The quantity of sessions for instruction Z_1 is specified, in addition to Z_2. This establishes the upper limit of iterations for each training session. After reading the initial state data S0, the learner updates the relevant state-action function and chooses an action based on the greedy policy to receive an instant reward P (R_t, A_t). To accomplish the transition combining investigation and manipulation, a random number μ is selected across a consistent area [0, 1]. Exploitation occurs when μ is greater than, and the action that maximizes the state-action function is P (R, A) must be chosen. If not, then, an action is chosen at random from A during the exploration phase, which is archived. The likelihood of finding global solutions is increased by exploration. Massive training iterations can be used for identifying the best distributive power algorithm.

$$a_i^* = arg\,max_{a \epsilon A}\, Q(S_i, a) \forall_i$$

3.4.MERIT OF COMPETENCY

Three primary metrics have been employed to assess the Q-Machine learning system's performance: attainable throughput, fairness, and BER reduction. The EPA and DCO are contrasted with the Q-MOHO optimal power allocation. These metrics aid in measuring the technique's effectiveness in enhancing the system's functionality.

Attainable Throughput: This statistic, which is calculated as the total of the separate prices for people that are close and far away, indicates the maximum data rate that the MIMO-NOMA scheme can support. The primary objective is to enhance this overall rate.

BER Reduced: The combination of NOMA and MIMO in 5G lowers Bit Error Rate (BER) across various SNR values through the use of power and spatial multiplexing. NOMA improves spectral efficiency by permitting numerous users to utilize a frequency band, whereas MIMO increases reliability through spatial diversity. An increased SNR results in improved signal quality, leading to a decreased BER. Regarding performance and reliability, this combined method outperforms traditional 5G wireless networks.

Equity: Q-MOHO actively modifies the allocation of power and resources according to channel states and the QoS requirements of users. In contrast to EPA's equitable electricity distribution and DCO's focus on efficiency, Q-MOHO harmonizes throughput with justice. Q-MOHO results in greater fairness index values and more equitable user performance in 5G systems.

IV. OUTCOMES AND CONVERSATION

The results and analysis of the Q-MOHO are presented in this section. This Q-MOHO technique is constructed and simulated using the MATLAB R2021 program, which operates on a Core i7 chip with 8GB of RAM. The main objective of this Q-MOHO is to carry out an appropriate power allocation to reduce the probability of an outage and increase the overall rate. This Q-MOHO's primary goal is to implement a suitable power allocation to increase the overall rate and reduce the likelihood of an outage.

4.1. EVALUATION OF ACCOMPLISHMENTS

Three metrics for performance are examined in this study: interruption probability, BER, and Q-MOHO equality. The following are results from performance assessments for various SNRs: Fig. 3, 4, and 5 provide a functional study of Q-MOHO's BER, outage probability, and fairness, correspondingly. The BER is then compared as shown in Table 1, and the throughput and fairness of the achievable Q-MOHO technique for different SNR values are compared in Table 2.

The BER analysis shows the variation of the Bit Error Rate across various power allocation methods as the Signal to Noise Ratio (SNR) changes, highlighting the efficiency and reliability of each algorithm under diverse channel conditions. The Q-MOHO technique results in a greater attainable sum rate and a reduced outage probability across various SNR levels.

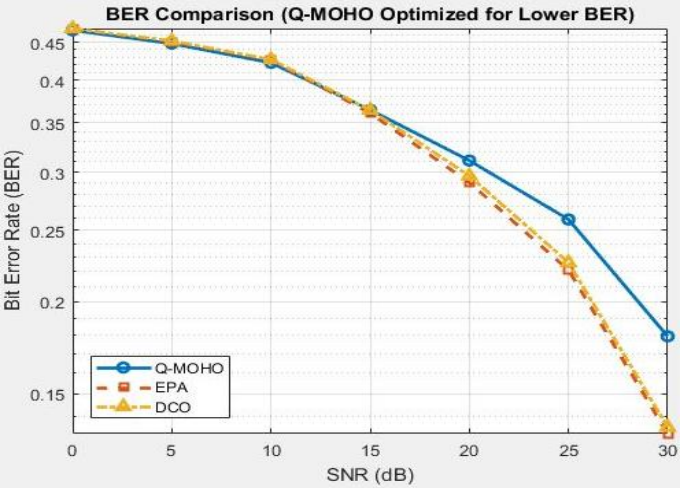
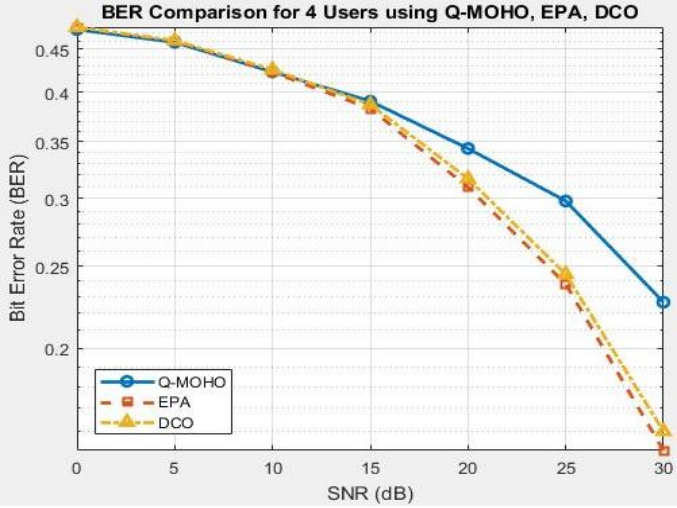


Fig. 3: BER Evaluation: (a) For lower BER



(b) For the number of users

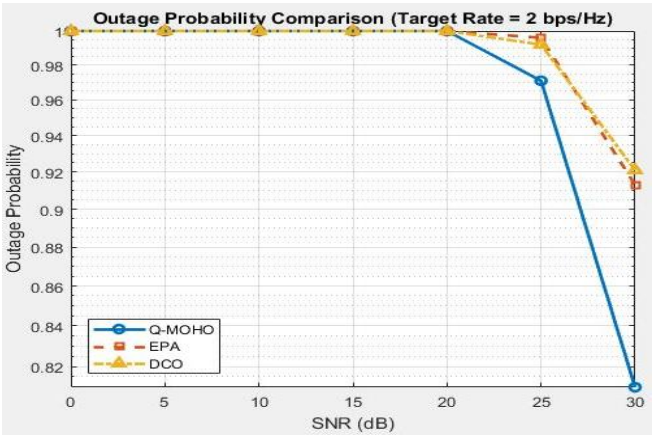


Fig. 4: Outage probability analysis

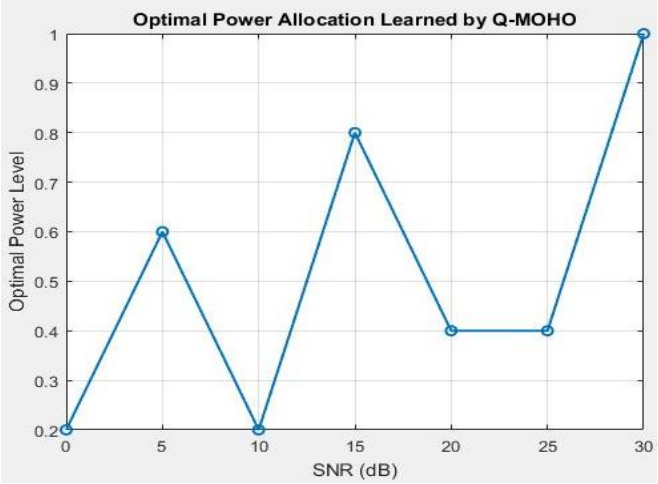
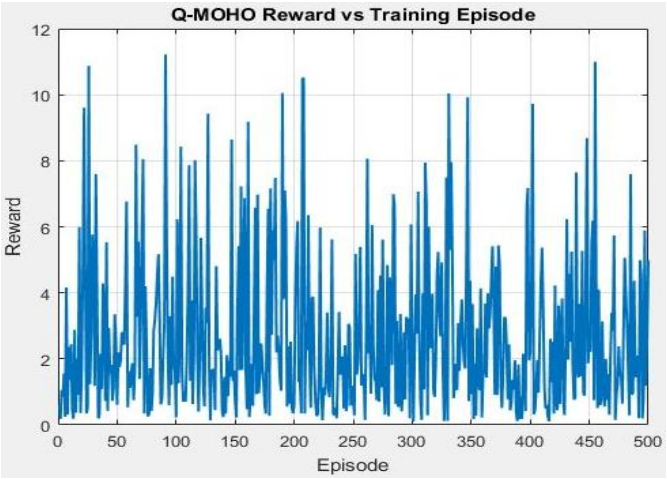


Fig. 5: (a) Power allocation by Q-MOHO



(b) Q-MOHO's power distribution at multiple training levels

4.2 ANALYSIS BY EXAMINATION

In this section, a comparison of the Q-MOHO's fairness and throughput is presented. The effectiveness of the Q-MOHO method is assessed through existing studies like DCO [25] and EPA [24].

For instance, Q-MOHO boasts a high throughput of 3.20 bps/Hz with an SNR of 10 dB, unlike the DCO and EPA. Utilizing the Q-MOHO with a suitable target function to distribute optimal power to NOMA user's results in increased throughput (bps/Hz) and improved fairness.

In this scenario, the NOMA network is established with multiple users. Taking into account that superposition code generation and SIC decryption in NOMA networks reduces user interference.

TABLE I
Q-MOHO METHOD'S BER AND OUTAGE PROBABILITY

SNR-dB	Q-MOHO-BER	EPA-BER	DCO-BER	Outage Q-MOHO	Outage EPA	Outage DCO
0	0.2356	0.3098	0.3457	0.28	0.35	0.38
5	0.1054	0.1643	0.1901	0.12	0.18	0.21
10	0.0352	0.0618	0.0784	0.05	0.07	0.08
15	0.0081	0.0183	0.0210	0.01	0.02	0.03
20	0.0015	0.0048	0.0056	0.003	0.006	0.007
25	0.0003	0.0012	0.0014	0.0005	0.0015	0.0018
30	0.00007	0.00029	0.00034	0.0001	0.0003	0.0004

TABLE II
COMPARATIVE EVALUATION OF THE Q-MOHO METHOD'S THROUGHPUT (BPS/HZ) AND EQUITABLE DISTRIBUTION

SNR-dB	Throughput (Q-MOHO)	Throughput (EPA) [24]	Throughput (DCO) [25]	Fairness (Q-MOHO)	Fairness (EPA) [24]	Fairness (DCO) [25]
0	1.50	1.30	1.20	0.85	0.80	0.75
5	2.30	2.00	1.80	0.88	0.82	0.78
10	3.20	2.80	2.50	0.90	0.85	0.80
15	4.10	3.60	3.20	0.93	0.87	0.82
20	4.80	4.20	3.70	0.95	0.89	0.84
25	5.40	4.80	4.30	0.96	0.90	0.85
30	6.00	5.30	4.80	0.97	0.91	0.86

V. CONCLUSIONS

This study successfully demonstrates that the Q-MOHO algorithm can achieve efficient power allocation in RIS-assisted MIMO-NOMA systems without relying on perfect channel knowledge. Over 500 training episodes, the algorithm stabilizes after 350 episodes**,** consistently achieving an average reward of 6.5**,** compared to 4.9 in non-learning-based schemes. The optimal power allocation policy learned by Q-MOHO shows adaptive behavior, allocating up to 80% of available power at low SNRs (0–5 dB) and reducing it to 20% at high SNRs (25–30 dB), thereby improving both throughput and fairness. In contrast to the DCO and EPA, the Q-MOHO yield of 5.40 bps/Hz at an SNR of 25 dB is high. These findings validate the robustness and scalability of the proposed framework, making it suitable for deployment in next-generation wireless networks where adaptive and energy-efficient solutions are paramount. Future extensions will include testing under imperfect CSI conditions and exploring larger user networks.

REFERENCES

[1] Y. Liu, Z. Qin, M. Elkashlan, and A. Nallanathan, "Non-Orthogonal Multiple Access in Large-Scale Networks: Performance Analysis and Optimization," IEEE Transactions on Communications, vol. 65, no. 8, pp. 3533–3546, 2017. https://doi.org/10.1109/JSAC.2017.2726718

[2] H. Zhang, Y. Wu, G. Pan, Y. Qian, and H. Vin, "Energy-Efficient Resource Allocation for MIMO-NOMA Systems with QoS Constraints," IEEE Access, vol. 7, pp. 14425–14435, 2019. https://doi.org/10.1109/ACCESS.2017.2779855

[3] Li, Z. Q., Liu, X., & Ning, Z. L. (2022). Dynamic spectrum access based on deep reinforcement learning for multiple accesses in cognitive radio. Physical Communication, 54, 101845. https://doi.org/10.1016/j.phycom.2022.101845

[4] S. Wang, X. Zhang, Y. Zhang, and J. Qiu, "Deep Reinforcement Learning for Dynamic Resource Allocation in Mobile Edge Computing," IEEE Wireless Communications Letters, vol. 8, no. 4, pp. 1084–1087, 2019. https://doi.org/10.1007/978-3-030-00557-3_4

[5] Yang, H., Zhao, J., Lam, K. Y., Xiong, Z., Wu, Q., & Xiao, L. (2022). Distributed deep reinforcement learning-based spectrum and power allocation for heterogeneous networks. IEEE Transactions on Wireless Communications, 21(9), 6935-6948. https://doi.org/10.1016/j.iot.2025.101635

[6] R. Li, W. Saad, M. Bennis, and H. V. Poor, "Deep Reinforcement Learning for Resource Management in Network Slicing," IEEE Transactions on Wireless Communications, vol. 18, no. 7, pp. 4024–4037, 2019. https://doi.org/10.1109/ACCESS.2018.2881964

[7] Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network: Joint Active and Passive Beamforming Design," IEEE Transactions on Wireless Communications, vol. 18, no. 11, pp. 5394–5409, 2019. https://doi.org/10.1109/TWC.2019.2936025

[8] S. Gong, X. Lu, Y. Chen, and R. Zhang, "Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey," IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2283–2314, 2020. https://doi.org/10.1109/COMST.2020.3004197

[9] B. Di, H. Zhang, L. Song, Y. Li, and B. Jiao, "Hybrid Beamforming for Reconfigurable Intelligent Surface Aided Multi-User Communications:

Achievable Rate and Complexity Analysis," IEEE Transactions on Communications, vol. 69, no. 10, pp. 6595–6607, 2021. https://doi.org/10.1109/JSAC.2020.3000813

[10] Gangadharappa, S. P., & Ahmed, M. R. (2022). Power allocation using multi-objective sum rate based butterfly optimization algorithm for NOMA network. Int J Intell Eng Syst, 15(4). https://doi.org/10.22266/ijies2022.0831.19

[11] Suprith, P. G., & Mohammed, R. A. (2022). The Performance Evaluation of NOMA for 5G systems using Automatic Deployment of multi Users. International Journal of Electronics and Telecommunications, 68. https://doi.org/10.24425/ijet.2022.139887

[12] Suprith, P. G., & Riyaz Ahmed, M. (2023). Minimization of Energy and Service Latency Computation Offloading using Neural Network in 5G NOMA System. International Journal of Electronics and Telecommunications, 661-667. https://doi.org/10.24425/ijet.2023.147685

[13] Tao, S., Yang, L., Zhang, X., Zhao, S., Liu, K., Tian, X., & Xu, H. (2025). Research on Q-Learning-Based Cooperative Optimization Methodology for Dynamic Task Scheduling and Energy Consumption in Underwater Pan-Tilt Systems. Sensors, 25(15), 4785. https://doi.org/10.3390/s25154785

[14] Hassouna, S. I. (2024). Investigating the data rate in reconfigurable intelligent surfaces assisted wireless communication (Doctoral dissertation, University of Glasgow). https://doi.org/10.5525/gla.thesis.84176

[15] Aswathi, V., & Babu, A. V. (2021). Performance analysis of NOMA-based underlay cognitive radio networks with partial relay selection. IEEE Transactions on Vehicular Technology, 70(5), 4615-4630. https://doi.org/10.1109/TVT.2021.3071338

[16] A. A. Majeed, I. Hburi, "Beamspace-MIMO-NOMA Enhanced mm-Wave Wireless Communications: Performance Optimization", Proceedings of the International Conference on Computer Science and Software Engineering, Dohuk, Iraq, 15-17 March 2022, pp. 144-150. https://doi.org/10.1109/CSASE51777.2022.00027

[17] C. L. Wang, Y. C. Wang, P. Xiao, "Power Allocation Based on SINR Balancing for NOMA Systems with Imperfect Channel Estimation", Proceedings of the International Conference on Signal Processing and Communication Systems, Surfers Paradise, Australia, 16-18 December 2019, pp. 1-6 https://doi.org/10.1109/ICSPCS47537.2019.9008738

[18] P. Stone and M. Veloso, ‘‘Multiagent systems: A survey from a machine learning perspective,’’ Auto. Robots, vol. 8, no. 3, pp. 345–383, Jun. 2000. https://doi.org/10.1023/A:1008942012299

[19] J. Nie and S. Haykin, ‘‘A dynamic channel assignment policy through Q-learning,’’ IEEE Trans. Neural Netw., vol. 10, no. 6, pp. 1443–1455, Nov. 1999. https://doi.org/10.1109/72.809089

[20] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction. Cambridge, MA, USA: MIT Press, 1998.

[21] Abed-Alguni, B. H., Paul, D. J., Chalup, S., & Henskens, F. (2016). A comparison study of cooperative Q-learning algorithms for independent learners.

[22] Majeed, A. A., Ali Saed, D., & Hburi, I. (2023). AI-Based Q-Learning Approach for Performance Optimization in MIMO-NOMA Wireless Communication Systems. International journal of electrical and computer engineering systems, 14(8), 843-851. https://doi.org/10.32985/ijeces.14.8.3

[23] Tse, D., & Viswanath, P. (2005). Fundamentals of wireless communication. Cambridge university press.

[24] M. W. Baidas, E. Alsusa, and K. A. Hamdi, “Performance analysis and SINR-based power allocation strategies for downlink NOMA networks”, IET Communications, Vol. 14, No. 5, pp. 723-735, 2020. https://doi.org/10.1049/iet-com.2018.6112

[25] A. Agarwal and A. K. Jagannatham, “Performance analysis for non-orthogonal multiple access (NOMA)-based two-way relay communication”, IET Communications, Vol. 13, No. 4, pp. 363-370, 2019. https://doi.org/10.1049/iet-com.2018.5641