Maximizing DRL-based Energy Efficiency in IRS-NOMA using a DDPG Algorithm for the Next Generation of Wireless Communications

Kamil Audah
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Department of Electronics Technologies, Southern Technical University, Iraq | Department of Computer Technology Engineering, Faculty of Engineering, Iraq University College, Iraq
thegenusnabster@gmail.com (corresponding author)

Nor K. Noordin
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Wireless and Photonics Networks Research Center of Excellence (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Malaysia
nknordin@upm.edu.my (corresponding author)

Wala'a Hussein
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Department of Chemical Engineering and Petroleum Refining, Basrah University for Oil and Gas, Iraq | Department of Computer Technology Engineering, Faculty of Engineering, Iraq University College, Iraq
walaahussein613@gmail.com

Mod Fadlee B. A. Rasid
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Wireless and Photonics Networks Research Center of Excellence (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Malaysia
fadlee@upm.edu.my

Aduwati Sali
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Wireless and Photonics Networks Research Center of Excellence (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Malaysia
aduwati@upm.edu.my

Aymen Flah
National Engineering School of Gabes, University of Gabes, Tunisia | University of Business and Technology (UBT), College of Engineering, Jeddah, 21448, Saudi Arabia | MEU Research Unit, Middle East University, Amman, 11831, Jordan | Applied Science Research Center, Applied Science Private University, Amman, Jordan | The Private Higher School of Applied Sciences and Technologies of Gabes (ESSAT), University of Gabes, Gabes, Tunisia
flahaymening@yahoo.fr

Received: 18 April 2024 | Revised: 5 May 2024 | Accepted: 7 May 2024

 DOI: https://doi.org/10.48084/etasr.7536

ABSTRACT

Combining Intelligent Reflecting Surfaces (IRSs) with Non-Orthogonal Multiple Access (NOMA) effectively enhances communication. This study introduces a NOMA-assisted Downlink Transmission (DT) system, emphasizing Energy Efficiency (EE) optimization. EE, crucial in Wireless Communications (WCs), measures data transmission relative to energy consumption. This study focuses on a Deep Deterministic Policy Gradient (DDPG) algorithm that intelligently adjusts IRS phase-shift matrices and access point beamforming in NOMA DT. Beamforming directs signals to users for optimal strength and quality, while phase shift control enhances signal coverage and quality. Strategic IRS placement improves user signal transmissions. The simulation results demonstrate significantly improved EE compared to other algorithms, such as Deep Q Network (DQN) and Proximal Policy Optimization (PPO), showcasing the effectiveness of the combined IRS and NOMA approach in enhancing communication systems' EE.

Keywords-5G; deep deterministic policy gradient; DRL; energy efficiency; IRS-NOMA; optimization

I. INTRODUCTION

Advances in wireless technologies from 1G to 5G have been developed on the assumption that the wireless environment is uncontrollable and unalterable. Consequently, improvements have primarily focused on designing efficient transmission and reception technologies to compensate for these inherent challenges. However, the advances made at the communication endpoints are insufficient to fulfill the Quality of Service (QoS) requirements for networks beyond 5G. The concept of 6G involves a fundamental change in approach, where the wireless landscape is viewed as a dynamic entity capable of adaptation and optimization. This transformative approach can be achieved by integrating Intelligent Reflecting Surfaces (IRSs), a key technology for the development of 6G networks [1]. The 6G plan, which is designed to enable the connectivity of the expanding network of interconnected devices serving applications, such as smart homes, cities, and transportation, using IRSs, requires sophisticated resource allocation frameworks that prioritize Energy Efficiency (EE) and Spectral Efficiency (SE). IRSs have become a focal point because of their adept management of electromagnetic waves. These surfaces consist of integrated electronic circuits with programmable features that enable customized control over the propagation properties of the electromagnetic channels. This technology offers a cost-effective solution to expand EE and SE in Wireless Communication (WC) beyond the capabilities of 5G, unlocking diverse application possibilities.

The incorporation of IRSs can take various forms. An IRS can function as a passive reflector operating without Radio Frequency (RF) chains and requiring minimal power for tuneable chip manipulation [2]. A passive IRS offers benefits, such as scalable costs, lower energy consumption, and simplified interference management. Conversely, an active IRS incorporates RF chains and can function as an active holographic Multiple Input-Multiple Output (MIMO) system. This study signifies a progressive advancement beyond MIMO, offering the potential to leverage the propagation features of the electromagnetic channel with notable simplicity in practical implementations [3-4]. Furthermore, the characteristics of IRSs include easy integration into existing wireless networks and versatility for deployment in diverse structures, such as buildings, roadside billboards, windows, and indoor walls. Many studies have explored the benefits of integrating IRSs into WCs, considering aspects, involving minimized power transmission, reduced interference, and improved security [5]. Non-Orthogonal Multiple Access (NOMA) is positioned as a leader in the next-generation multiple access category in 6G communication networks, with anticipated advances in terrestrial, aerial, and maritime communications, providing faster, more efficient, and reliable connectivity [6]. The 6G network is designed to accommodate extensive connectivity with exceptionally low latency requirements, catering to applications, like smart homes, cities, and transportation. NOMA plays a pivotal role in fulfilling the ambitious goals of 6G, attracting both academic and industrial interest. It is recognized for its superior SE, providing fair network access and equitable treatment to users [7], and effectively addressing resource conflicts that arise from intense transmissions within limited resources. The combination of NOMA with IRSs offers a compelling opportunity to enhance massive connectivity while improving both SE and EE [8-9]. Notably, NOMA networks empowered by IRSs exhibit enhanced adaptability in contrast to conventional NOMA networks. The key advantage of IRSs lies in their ability to intelligently adjust the direction of received signals (channel vectors), simplifying NOMA deployment [10]. Acting as a passive amplifier and forwarder, an IRS improves signal transmission and is particularly valuable in cell-free massive MIMO systems [11]. IRSs offer superior EE compared to traditional AF relays [12-13] and have been proven beneficial for multicast communications [14-15], demonstrating considerable improvements in the EE of IRS-assisted broadcast communication systems.

In [16], the balance between EE and SE was explored in IRSs of several mobile units in an uplink system to optimize phase shift and covariance matrix. In [17], the challenge of maximizing security in IRS-assisted multicast communication was addressed by effectively suppressing unwanted signals. However, the specific considerations related to EE in IRS-assisted multicast communications present an open area for exploration [18- 19]. Improving EE and SE is crucial in the WC dynamic landscape [20-21]. Combining IRSs and NOMA presents novel opportunities to address NOMA's constraints by leveraging IRSs' wave manipulation and NOMA's multiplexing [22-23]. Optimizing EE is crucial, with IRSs offering cost-effective implementation and the potential for enhanced system performance [12, 24]. Recent studies focused on various IRS-enabled aspects, including channel modeling, estimation, encoding, modulation, and performance evaluation [25-26]. This study aims to optimize EE in IRS-NOMA using a Deep Reinforcement Learning (DRL) algorithm to minimize power consumption.

A. Research Motivation

The next generation WC demands high data rates with limited energy. This study explores DRL to optimize IRS-NOMA systems. The Deep Deterministic Policy Gradient (DDPG) algorithm helps the IRS dynamically adjust signals, maximizing energy efficiency while enabling high data rates for future communication networks.

B. Research Contribution

This study significantly advances WC systems by improving EE in the IRS-NOMA context and presenting an optimization algorithm based on DRL principles, particularly enhancing EE in NOMA-assisted WC systems using the DDPG algorithm.

- This study explores the integration of IRS and NOMA into DT WC systems. This integration is studied with a focus on optimizing the EE within the system. This study considered the direct link of transmission between BS and users. The purpose of combining the IRS and NOMA was to improve the efficiency of data transmission while minimizing energy consumption.
- This study employs DDPG, a DRL algorithm, to optimize the beamforming and IRS phases for maximum EE in communication systems. This approach allows for adaptation to changing conditions and enhanced system performance.
- This study evaluates DDPG's performance (reward, loss, exploration) during training to assess its learning and adaptation. Simulations compare DDPG with other algorithms (DQN, FL, PPO) to demonstrate its effectiveness in optimizing beamforming for EE in WC systems.

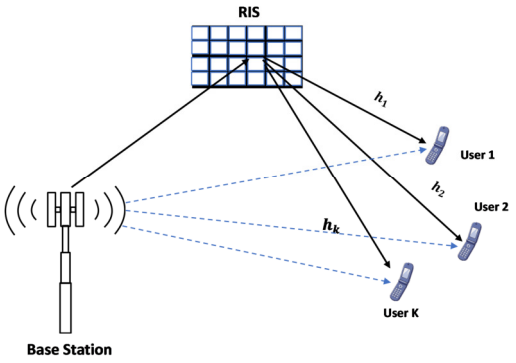

Fig. 1. The proposed IRS-NOMA system model.

II. SYSTEM MODEL

A. IRS-NOMA System

The IRS-NOMA communication system comprises an IRS, a BS, and multiple users, as observed in Figure 1. The BS is equipped with antennas denoted by $\mathcal{M} = \{1, 2, \dots, M\}$ and $\mathcal{K} = \{1, 2, \dots, K\}$. The IRS comprises N reflection units, represented by $\mathcal{N} = \{1,2, \dots, N\}$. The IRS can independently reflect incoming signals according to CSI and modify the amplitude or phase of each reflection unit to manage the signal's directions. The signal transmitted in the IRS-NOMA DT communication system can be expressed as:

$$s = \sum_{i=1}^{K} \omega_i x_i \quad (1)$$

In this context, $\omega_i \in \mathbb{C}^{\mathrm{A}}$ is the generated precoding matrix and $x_i \in \mathbb{C}^{\mathrm{A}}$ is the transmitted signal [27]. The signal received by the users can be characterized by:

$$y_k = (H_1 \Theta\, G + H_2)s + e_k \quad (2)$$

The communication channel parameters between the BS and the user, BS and IRS, and IRS and the user are denoted as $H_2 \in \mathbb{C}^{M\times 1}$, $G \in \mathbb{C}^{M\times N}$, and $H_1 \in \mathbb{C}^{N\times 1}$, respectively. The IRS is represented by a diagonal phase-shift matrix denoted by $\Theta = \mathrm{diag}\left(e^{j\theta_1}, \dots, e^{j\theta_N}\right) \in \mathbb{C}^{N\times N}$, where $\theta_n \in [0, 2\pi]$, $n \in [1, \dots, N]$, and $e^k \sim \mathcal{CN}(0, \sigma^2)$. The additive white Gaussian noise with variance σ^2 is present in the system to facilitate NOMA transmission. All the users are grouped based on channel quality compared to the BS. NOMA technology relies on Successive Interference Cancellation (SIC). Users decode the strongest signal first and then remove its interference from the received signal to decode weaker signals. This means that stronger signals act like noise for weaker users. The equation representing the received signal by the user is given by:

$$y_k = (H_1 \Theta\, G + H_2)s' + e_k \quad (3)$$

where $s' = \sum_{i=1}^{K} \omega_i x_i$. Successful decoding of the ordered signal facilitates the determination of the transmission rate for user k employing the following equation:

$$\mathcal{R}_k = \log_2\left(\frac{|(\mathrm{H}_1\Theta\,\mathrm{G}+\mathrm{H}_2)\omega_k|^2}{\left|(\mathrm{H}_1\Theta\,\mathrm{G}+\mathrm{H}_2)\sum_{j=k+1}^{K}\omega_k\right|^2 + e_k^2}\right) \quad (4)$$

In this context, the expression:

$$\left(\frac{|(\mathrm{H}_1\Theta\,\mathrm{G}+\mathrm{H}_2)\omega_k|^2}{\left|(\mathrm{H}_1\Theta\,\mathrm{G}+\mathrm{H}_2)\sum_{j=k+1}^{K}\omega_k\right|^2 + e_k^2}\right)$$

represents a signal-to-interference-ratio (SINR) at the k^{th} user, which leads to formulating the system's SE as follows:

$$\mathcal{R} = \sum_{k=1}^{K} \mathcal{R}_k \quad (5)$$

B. Optimization Problem

To improve the IRS-NOMA system, EE was determined using its SE ratio to its total power consumption, as defined by:

$$max_{\Phi G}\eta(G(t), \Phi(t), H_1, H_2) = \frac{\mathcal{R}}{\mu\sum_{k=1}^{K} p_k + P_{BS} + KP_u + NP_n} \quad (6)$$

$$s.t\ tr\{GG^{\mathcal{H}}\} \le P_t$$

$$|\varphi_n| = 1 \quad \forall n = 1,2, \dots, N$$

In this expression, $\mu\sum_{k=1}^{K} p_k, P_{BS}, P_n, P_U$, are the combined power consumption of the transmitting BS, the power dissipation at the BS, the power loss at the mobile UT, and the power consumption of the IRS, respectively [28].

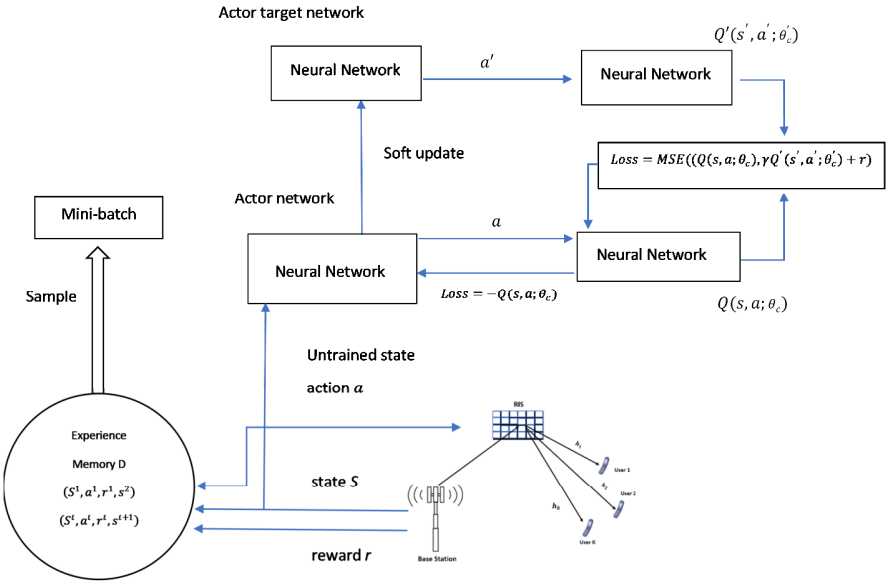

Fig. 2. Schematic diagram of the optimization model for the DT system incorporating IRS-NOMA and using the DDPG algorithm.

III. DEEP REINFORCEMENT (DRL) ALGORITHM

A. Deep Reinforcement Learning (DRL)

DRL interacts with environments via Markov Decision Processes (MDPs), executing actions based on observed states and receiving rewards. The goal is to discover optimal strategies for maximizing cumulative rewards. MDP policies depend on current states, represented by $W = [S, A, P, R, \gamma]$, where $S = \{s_1, s_2, \dots, s_n\}$ denotes the sets of states (S), $A = \{a_1, a_2, \dots, a_n\}$ denotes the sets of actions, $P^a_{s \to s'} = P_r(s'|(s,a))$ denotes the probability of transitioning from the current S to the subsequent S' when A is performed, $R(s,a) = \mathbb{E}[R_{t+1}|s,a]$ denotes the instant reward obtained when the agent performs A in the current S, and γ is the discount factor that determines the significance of each reward as it varies. The cumulative reward of the agent is the total of the rewards gathered throughout a series of interactions with the environment:

$$G(s) = \sum_{t=0}^{\infty} \gamma^t R(s_t)\ 0 \le \gamma < 1 \quad (7)$$

The optimization of the agent is to find a strategy $\pi(a|s)$ that maximizes the overall reward $G(s)$, where $\pi(a|s) = P(a|s)$, representing the likelihood of taking A in the current S:

$$V_\pi(s) = \mathbb{E}_\pi[R(s,a) + \gamma\, V_\pi(S_{t+1}|S_t = s)] \quad (8)$$

Given the S at the time t, the resultant function represents the anticipated reward. The state-action pair function is the probable reward of an S at t following the selection of A, and it is calculated as follows:

$$Q_\pi(s,a) = \mathbb{E}[R_{t+1} + \gamma Q_\pi(S_{t+1}, A_{t+1})|S_t = s, A_t = a] \quad (9)$$

The optimal solution for the following equations is determined using the Bellman optimality criterion as in [22]:

$$V_*(s) = max_\pi V_\pi(s) = max_a \left(R(s,a) + \gamma \sum_{s' \in S} P^a_{s \to s'} V_*(s')\right) \quad (10)$$

$$Q_*(s,a) = max_\pi Q_\pi(s,a) = R(s,a) + \gamma \sum_{s' \in S} P^a_{s \to s'} max_{a'} Q_*(s', a') \quad (11)$$

In this context, $V_*(s)$ is the value obtained from choosing a, representing the optimal value by evaluating every conceivable outcome in the current S. $Q_*(s,a)$ is the persistent value obtained by considering all potential S resulting from each A and selecting the best A.

B. Deep Deterministic Policy Gradient (DDPG) Algorithm

This study aims to enhance EE in the IRS-NOMA system by considering dynamic variations in wireless cellular network parameters. Figure 2 illustrates the optimization model for the IRS-NOMA DT system using the DDPG algorithm, segmented into actor and critic components with symmetric characteristics. A Deep Neural Network (DNN) with three layers facilitates generalization.

$$\theta_a^* = argmax Q_{\theta_a}(s, a; \theta_c | a = A(s, \theta_a)) \quad (12)$$

$$\theta_c^* = \text{argmin}_{\theta_c} \frac{1}{2} (Q(s,a;\theta_c)\big|_{a=A(s,\theta_a)} - R(s,a))^2 \quad (13)$$

Both θ_a^* and θ_c^* are differentiable. Using the chain rule, the gradients of these variables can be derived as:

$$\nabla \theta_a = \nabla_a Q(s_c, a; \theta_c)|_{a=A(s,\theta_a)} \nabla_{\theta_a} A(s; \theta_a) \quad (14)$$

$$\nabla \theta_c = (Q(s_c, a; \theta_c) - R(s,a)) \nabla_{\theta_c} Q(s,a;\theta_c)\big|_{a=A(s,\theta_a)} \quad (15)$$

Software updates were implemented to modify parameters in the target network. This was accomplished using an initial update rate that progressed gradually. The equations governing this process are:

$$\theta_a' = \tau \theta_a + (1 - \tau) \theta'_{a'} \quad (16)$$

$$\theta_c' = \tau\theta_c + (1-\tau)\theta'_{c'} \quad (17)$$

where τ is the update coefficient.

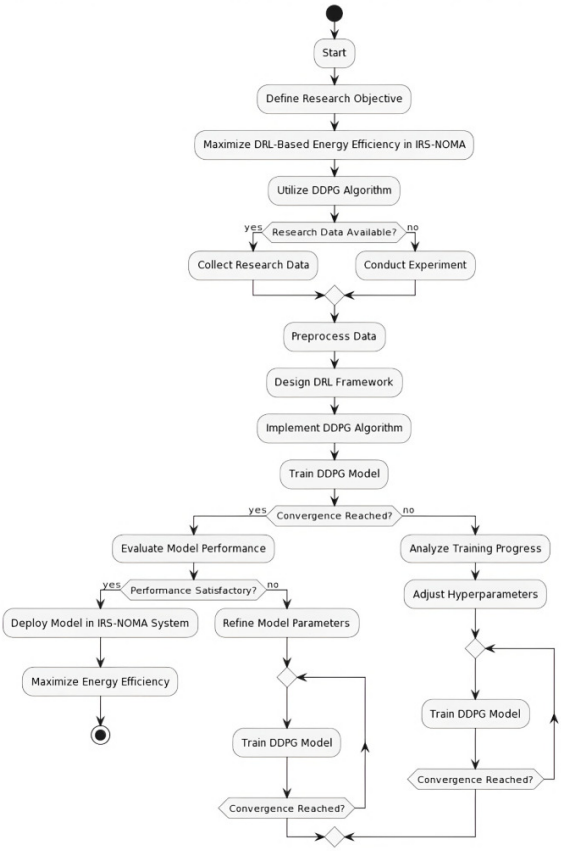

Fig. 3. Flowchart of DRL optimization algorithm.

This study utilized the IRS-NOMA system's EE, which was improved deploying the DDPG algorithm. Taking advantage of the DDPG algorithm's ability to collect real-time channel status information, the current state, possible actions, and immediate rewards were defined:

- State: In the t^{th} step, the state $s^{(t)}$ is collectively determined by the channel matrices H_1 and H_2, the received power of users, and the action taken in the $(t\text{-}1)^{\text{th}}$ step. Due to the complexity of the matrices in the system, employing absolute operators to calculate the transmitted and received powers results in loss of information. Therefore, it is essential to differentiate between the imaginary part and the real part of the signal that has been transmitted.
- Action: The action is generated by combining the transmit beamforming matrix G and the phase shift matrix Φ. Similarly, the real and imaginary components of G are split into two parts, $\mathrm{Re}\{G\} + \mathrm{Im}\{G\}$, and Φ is separated into its real and imaginary parts as $\mathrm{Re}\{\Phi\} + \mathrm{Im}\{\Phi\}$.
- Reward: In DRL, in the t^{th} step, the reward is calculated as the EE, represented by $\eta(\mathrm{G}(t), \Phi(t), H_1, H_2)$ in the context of IRS-NOMA WCs.

The DDPG algorithm employs four neural networks. It is assumed that each network consists of L layers, with each layer containing M_l neurons. The complexity of neurons varies depending on their type [29]. The sum of nodes for batch normalization, ReLU, and tanh layers are denoted as M_b, M_r, and M_t, respectively. M_a are the nodes of the actor network (η_a) and M_c are the nodes of the critic network (η_c). The requirement for floating point calculations in each batch normalization node, ReLU node, and tanh node is 5, 1, and 6, respectively [30]. Both the η_a and η_c collaborate during the training process. Therefore, their complexities are:

$$\mathcal{O}(5M_b^c + M_r^c + 6M_t^c + \sum_{L=0}^{L-1} M_l^c + M_{l+1}^c) \text{ and}$$

$$\mathcal{O}(5M_b^a + M_r^a + 6M_t^a + \sum_{L=0}^{L-1} M_l^a + M_{l+1}^a)$$

Meanwhile, the target network's complexity is

$$\mathcal{O}(\sum_{L=0}^{L-1} M_l^c + M_{l+1}^c) + \mathcal{O}(\sum_{L=0}^{L-1} M_l^a + M_{l+1}^a)$$

Therefore, the overall complexity of the algorithm is:

$$\mathcal{O}(N.T.\big((5M_b^c + M_r^c + 6M_t^c + \sum_{L=0}^{L-1} 2M_l^c + M_{l+1}^c) + (5M_b^a + M_r^a + 6M_t^a + \sum_{L=0}^{L-1} 2M_l^a + M_{l+1}^a)\big)) \quad (18)$$

The number of training episodes and the number of steps per training episode are denoted by N and T. Algorithm 1 presents the DDPG algorithm's pseudocode in the context of an IRS-NOMA [31].

ALGORITHM 1: IRS-NOMA DDPG ALGORITHM

```
Input: The episode number, the learning rates for η_a and η_c, and the channel matrices H_1 and H_2
Set up a size D memory for an experience, establish parameters for the training of the η_a (θ_a), initialize parameters for the target η_a as equal to the η_a θ_a' = θ_a, define parameters for training the η_c θ_c, set parameters for the target η_c to match those of the η_c θ_c' = θ_c, prepare the matrix (G) for beamforming transmit, and matrix (Φ) for the phase shift.
For n = 1 to N do
  Compile and pre-process, H_1 and H_2 to obtain the first state s^1
  For t = 1 to T do
    Obtain the action a^t from the η_a
    Perform the action (a^t) and observe the immediate reward (r^t)
    Observe the succeeding state (s^{t+1})
    Store the experience parameters (s^t, a^t , r^t, s^{t+1} ) in the replay memory
    Calculate the critical gradient (∇θ_c) using (15) and update parameter (θ_c) using θ_c ← θ_c − η_c∇θ_c
    Calculate the gradient of the actor (∇θ_a) using (14) and adjust the parameter (θ_a) using θ_a ← θ_a − η_a∇θ_a
    Update the s^t = s^{t+1} states in η_a and η_c
  End For
End For
Output: the function value of Q and the action
```

IV. SIMULATION RESULTS AND ANALYSIS

The main objective of the simulation was to assess and maximize the EE and then enhance the performance of the DDPG algorithm in a system. The simulation considered a small-scale Rayleigh fading, a standard model in WC that mimics signal strength variations due to factors such as multipath propagation. Table I displays the settings and parameters of the simulation. TensorFlow served as the basis for the simulation environment, the dataset was simulated using MATLAB 5G Simulator, and the proposed algorithm was implemented utilizing MATLAB and Python.

TABLE I. SIMULATION PARAMETERS

Parameter	Value	Description
θ_a	0.001	Learning rate for the η_a update
θ_c	0.001	Learning rate for the η_c update
D	100000	Experience replays buffer size
N	1000	The training episode number
T	10000	Number of steps in the training episode
P_{BS}	10 dBW	Power dissipation in the BS circuit
μ	1.2	Coefficients for power dissipation in the BS circuit
P_{UE}	10 dBW	Coefficients for power dissipation in the BS circuit
P_n	10 dBW	Power dissipation at the n^{th} elements of the IRS
K	2-10 user	Number of users served by the BS
M	100	Number of elements in the IRS that manipulate the signal
N_t	2	Number of transmit antennas at the BS
N_r	1	Number of receive antennas at each user.
P_t	15 -30 dBm	Total power transmitted by the BS
Distance BS to IRS	100 m	The distance between IRS and BS
Distance IRS to user	110 m	The distance between IRS and user
Path loss exponent (α)	2	This models the signal attenuation with distance
Fading	Exponential, Rayleigh	Overall signal weakening due to long distances
Doppler frequency	40 - 200 Hz	Define the Doppler frequency
Episodes	100 - 400	The number of training iterations.

Figure 4 depicts the training diagram for the DDPG algorithm, showing an increase in reward with training steps until stabilization. More users decrease EE when the transmission power is constant. An increase in the transmitting power of the BS results in a reduction in EE. Furthermore, the overall EE of the system diminishes with an increase in the number of reflecting elements. The number of reflecting elements on the IRS can be carefully increased to achieve optimal performance. Figure 5 indicates a proportional increase in EE with increased BS transmission power and reflects a decrease in total EE with more reflecting elements. Adding reflecting elements benefits IRS performance. Figure 6 compares the DDPG algorithm in IRS-NOMA and CNN IRS-NOMA transmission, highlighting a growing disparity in achievable rate with an increased number of reflecting elements. The DDPG algorithm provides the best achievable rate compared to other algorithms.

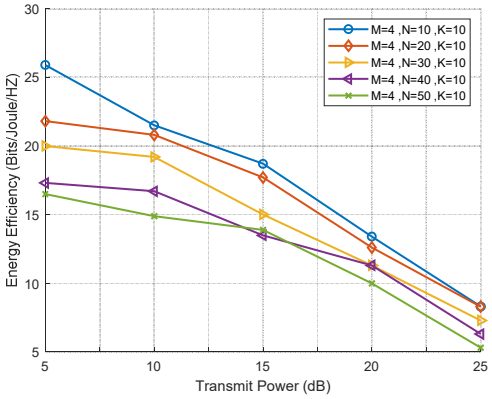

Fig. 4. Changes in EE in response to variations in the power transmitted by the BS for $N = 10, 20, 30, 40, 50$.

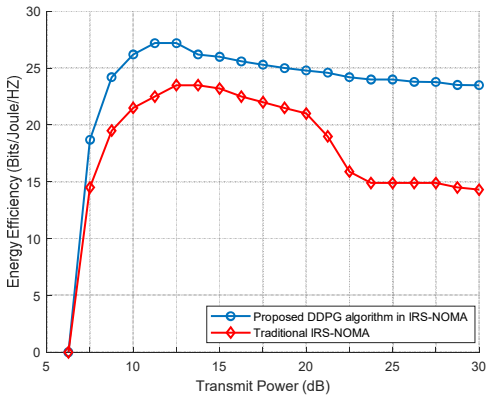

Fig. 5. Maximum EE reached using the DDPG algorithm.

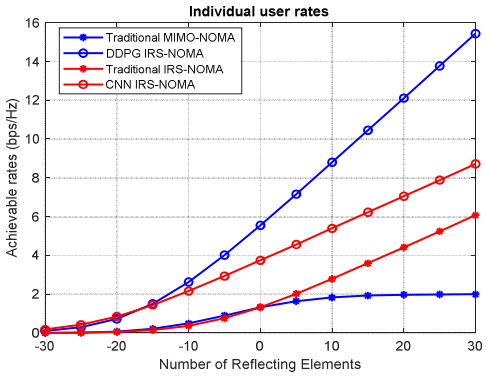

Fig. 6. Achievable rates against the number of reflecting elements for the DDPG algorithm against CNN and traditional IRS-NOMA.

NOMA technology proved to be beneficial for EE enhancement. In Figure 7, DDPG outperformed FL and IRS with random phases. Comparing RL algorithms, such as DDPG, with techniques, such as FL for EE in IRS-NOMA, is complex and context-dependent, suitable for continuous action spaces and complex policies. DDPG's performance relative to other methods depends on factors such as problem specifics and data. DDPG excels in continuous action spaces, offering fine-grained control in persistent problems. In comparisons (Figures 8 and 9), DDPG exhibits high EE in IRS-NOMA

setups, aligning closely with optimal EE. Utilizing DNN for policy and value function approximation enables DDPG to capture complex relationships and enhance performance.

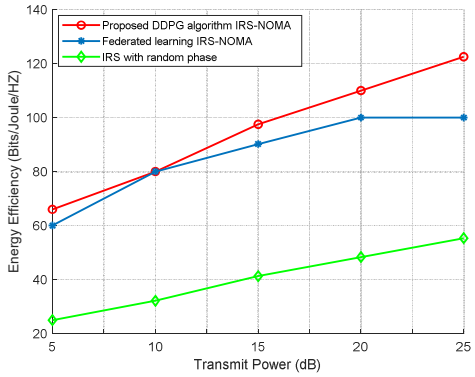

Fig. 7. Comparison of the effect of transmit power on the EE of an IRS-NOMA system and the proposed DDPG, Federated Learning IRS-NOMA [12], and random phase-based-IRS [30].

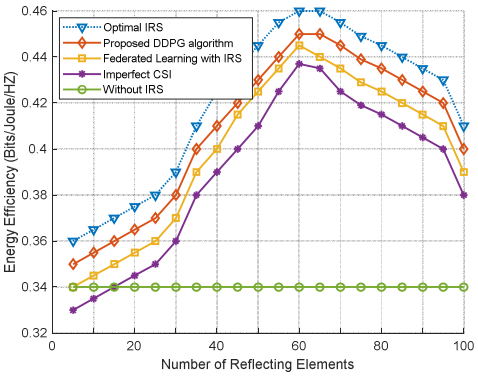

Fig. 8. Maximum achievable EE against the number of reflecting elements in an optimal IRS, the proposed DDPG algorithm, FL [28], imperfect CSI [29], and without IRS.

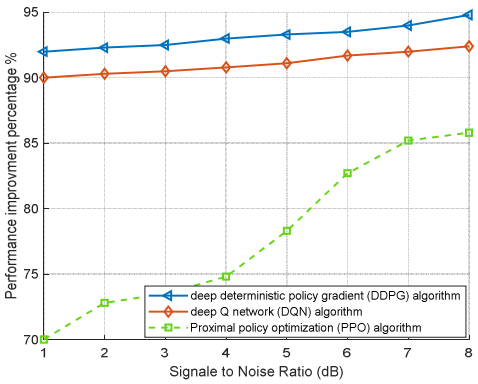

Fig. 9. Comparison of the improvements in performance against the signal-to-noise ratio of different optimization algorithms, namely, the proposed DDPG, PPO [32], and DQN [29].

Figure 10 discloses that the DDPG algorithm provides the minimum power consumption compared to traditional IRS-NOMA due to the optimization method followed, and this result can provide high EE compared to the traditional IRS method. In Figure 11, the proposed DDPG algorithm was able to provide a very high reward signal because it is an RL algorithm designed to maximize cumulative rewards in tasks involving sequential decision-making. DDPG can achieve high rewards in various environments. As a policy gradient algorithm, DDPG optimizes policy to maximize rewards, which is beneficial for finding optimal approaches in control problems, such as EE optimization in IRS-NOMA systems.

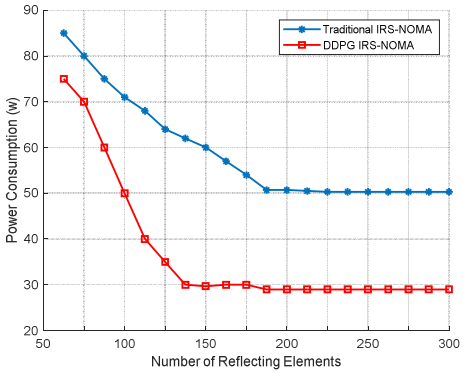

Fig. 10. Power consumption against the number of reflecting elements for DDPG and traditional IRS-NOMA.

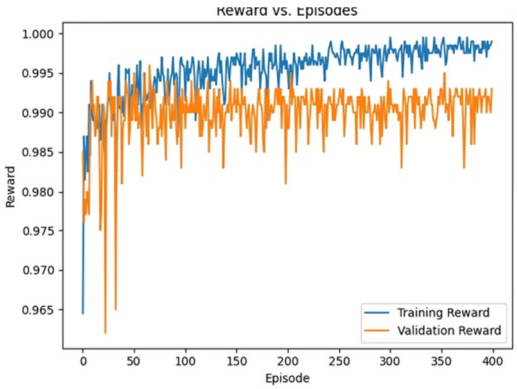

Fig. 11. Moving average reward concerning the episodes in the proposed DDPG algorithm for training and reward validation.

The design of the reward function significantly affects DDPG's performance. A well-designed function efficiently guides the agent to high rewards, which is beneficial in WCs, autonomous systems, and similar domains. Rewards suggest agent performance, with higher values indicating better performance. Figure 12 compares the EE of different IRS-assisted systems (DDPG-NOMA, FL-NOMA, Traditional) for varying SE levels, showing that as SE increases, EE reaches a maximum point. In particular, DDPG-NOMA achieves significantly higher EE compared to FL-NOMA and traditional methods at all SE levels. Figure 13 offers a focused view of the average performance of the DDPG algorithm over time, smoothing out variations in episode rewards. Monitoring this average reward is crucial to assess learning progress and algorithm convergence. Figure 13 provides valuable insight into the algorithm's learning and performance changes during training, serving as standard tools for evaluating training progress in RL tasks.

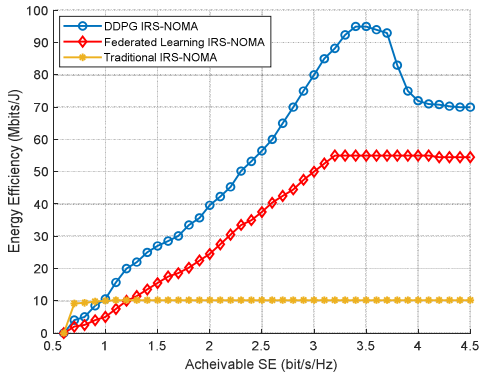

Fig. 12. Impact of SE on EE in IRS-NOMA systems with DDPG, FL, and traditional algorithms.

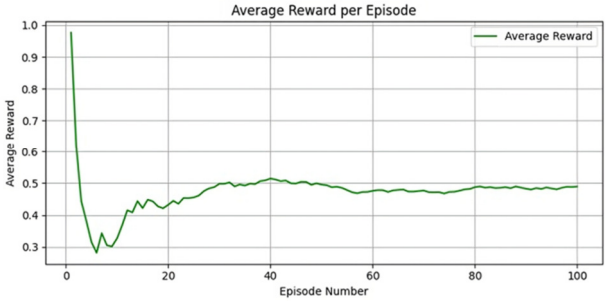

Fig. 13. Comparison of the average reward concerning episode numbers in the proposed DDPG algorithm.

η_c represents the Q values estimated by the DDPG algorithm, which are a fundamental concept in RL representing the expected cumulative reward that can be achieved by taking a specific action in a particular state. It is essential to monitor the Q values to understand the agent's assessment of the value of different actions in different states. Figure 14 provides critical insights into the learning progress of the DDPG algorithm through its η_c. Accurate Q value estimation is crucial for effective decision-making. The plot tracks how these estimates evolve over episodes, ideally converging to accurate values. Convergence indicates learning success. Figure 14 assesses the DDPG's algorithm learning progress and the quality of its Q value estimates, which are essential for informed decision-making in complex environments [33].

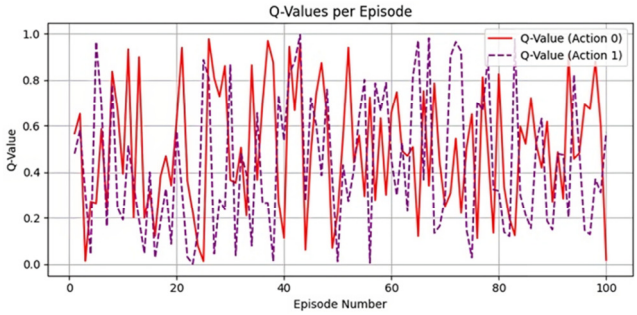

Fig. 14. Proposed DDPG algorithm's estimation of Q against the episode number using its η_c.

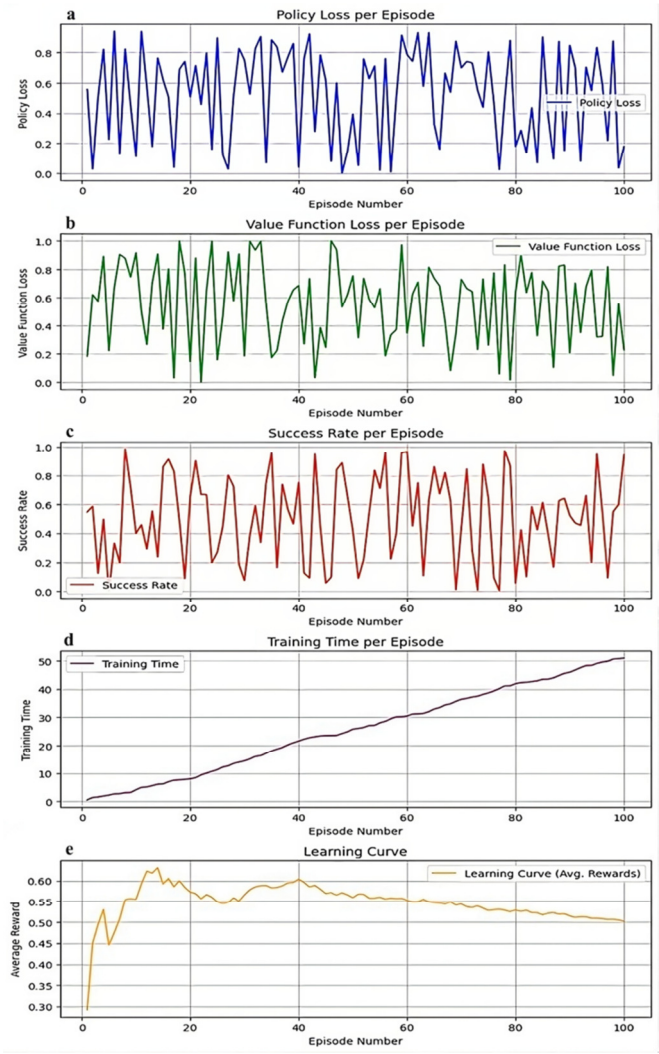

Fig. 15. Proposed DDPG algorithm's (a) policy loss, (b) value function loss, (c) success rate, (d) training time, and (e) learning curve.

Figure 15(a) depicts policy loss, a key RL metric that indicates the disparity between the predicted and optimal action probabilities. It guides DDPG's policy network learning, with decreased loss signaling policy improvement. Increasing policy loss can imply exploration or action uncertainty. Value function loss measures η_c accuracy in predicting the expected rewards. Decreasing loss signifies improved value estimation, crucial for action quality assessment and policy enhancement. The success rate measures the agent's achievement of the predefined goals to assess task performance. Training time measures the completion time of the DDPG algorithm, which is vital in resource-constrained settings as it indicates how long training for acceptable performance takes. The learning curve depicts performance changes over time, showing the agent's learning progress. Assessing these components offers insights into the performance of the DDPG algorithm. Figure 15 allows various aspects of the performance of the proposed DDPG algorithm to be evaluated, including policy improvement, the accuracy of the value function estimation, the agent's success in achieving specific goals, the time required for training, and

overall learning progress. These metrics and the learning curve provide valuable insights into the behavior and effectiveness of the algorithm in the given environment.

V. CONCLUSION AND RECOMMENDATIONS FOR FUTURE STUDIES

The DDPG algorithm, widely deployed in various domains, including WCs, stands out in IRS and NOMA systems. It optimizes transmit power levels for both IRS and NOMA users, to maximize EE while maintaining performance. This study improves EE in IRS-NOMA systems by refining power allocation and resource management, using the DDPG algorithm. This framework optimizes beamforming vectors and IRS phase shift, resulting in significant EE improvements. The simulation results demonstrated the superiority of DDPG over other algorithms, such as DQN and FL. Integration of NOMA and IRS proves effective in bolstering EE. Future studies can consider incorporating context-aware DRL with transfer learning to ameliorate efficiency by adapting to network scenarios and leveraging knowledge across them, utilizing contextual information (user traffic, channel conditions, etc.) for better DRL performance, and leveraging transfer learning to enhance efficiency across different network scenarios.

ACKNOWLEDGMENT

The authors thank the University Putra Malaysia, especially the Department of Computer and Communications System Engineering and WiPNET.

REFERENCES

[1] R. Alghamdi *et al.*, "Intelligent Surfaces for 6G Wireless Networks: A Survey of Optimization and Performance Analysis Techniques," *IEEE Access*, vol. 8, pp. 202795–202818, 2020, https://doi.org/10.1109/ACCESS.2020.3031959.

[2] Q. Wu and R. Zhang, "Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, Jan. 2020, https://doi.org/10.1109/MCOM.001.1900107.

[3] C. Huang *et al.*, "Holographic MIMO Surfaces for 6G Wireless Networks: Opportunities, Challenges, and Trends," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 118–125, Oct. 2020, https://doi.org/10.1109/MWC.001.1900534.

[4] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource Allocation for IRS-Assisted Full-Duplex Cognitive Radio Systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7376–7394, Sep. 2020, https://doi.org/10.1109/TCOMM.2020.3020838.

[5] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and Secure Wireless Communications via Intelligent Reflecting Surfaces," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2637–2652, Aug. 2020, https://doi.org/10.1109/JSAC.2020.3007043.

[6] Y. Liu *et al.*, "Evolution of NOMA Toward Next Generation Multiple Access (NGMA) for 6G," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1037–1071, Apr. 2022, https://doi.org/10.1109/JSAC.2022.3145234.

[7] A. Ihsan, W. Chen, S. Zhang, and S. Xu, "Energy-Efficient NOMA Multicasting System for Beyond 5G Cellular V2X Communications With Imperfect CSI," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10721–10735, Dec. 2022, https://doi.org/10.1109/TITS.2021.3095437.

[8] Z. Ding and H. V. Poor, "A Simple Design of IRS-NOMA Transmission," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1119–1123, Feb. 2020, https://doi.org/10.1109/LCOMM.2020.2974196.

[9] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent Reflecting Surface Aided Multigroup Multicast MISO Communication Systems," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3236–3251, 2020, https://doi.org/10.1109/TSP.2020.2990098.

[10] C. Pan *et al.*, "Multicell MIMO Communications Relying on Intelligent Reflecting Surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, Dec. 2020, https://doi.org/10.1109/TWC.2020.2990766.

[11] Y. Wang, H. Lu, D. Zhao, and H. Sun, "Energy Efficiency Optimization in IRS-Enhanced mmWave Systems with Lens Antenna Array," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, Dec. 2020, pp. 1–6, https://doi.org/10.1109/GLOBECOM42002.2020.9348266.

[12] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy Efficiency and Spectral Efficiency Tradeoff in RIS-Aided Multiuser MIMO Uplink Transmission," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1407–1421, 2021, https://doi.org/10.1109/TSP.2020.3047474.

[13] L. Du, C. Huang, W. Guo, J. Ma, X. Ma, and Y. Tang, "Reconfigurable Intelligent Surfaces Assisted Secure Multicast Communications," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1673–1676, Jul. 2020, https://doi.org/10.1109/LWC.2020.3001119.

[14] C. Huang, R. Mo, and C. Yuen, "Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, Dec. 2020, https://doi.org/10.1109/JSAC.2020.3000835.

[15] Q. Wu and R. Zhang, "Beamforming Optimization for Wireless Network Aided by Intelligent Reflecting Surface With Discrete Phase Shifts," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020, https://doi.org/10.1109/TCOMM.2019.2958916.

[16] H. Ghavami and B. Akhbari, "Secrecy performance analysis of IRS-NOMA systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, Jul. 2023, Art. no. 57, https://doi.org/10.1186/s13638-023-02266-z.

[17] S. P. Le, H. N. Nguyen, N. T. Nguyen, C. H. Van, A. T. Le, and M. Voznak, "Physical layer security analysis of IRS-based downlink and uplink NOMA networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, Oct. 2023, Art. no. 105, https://doi.org/10.1186/s13638-023-02309-5.

[18] B. C. Nguyen, T. Manh Hoang, A. T. Le, V. D. Nguyen, and P. T. Tran, "Performance analysis of intelligent reflecting surface aided full-duplex amplify-and-forward relay networks," *International Journal of Communication Systems*, vol. 35, no. 10, 2022, Art. no. e5172, https://doi.org/10.1002/dac.5172.

[19] N. D. Nguyen, A. T. Le, M. Munochiveyi, F. Afghah, and E. Pallis, "Intelligent Reflecting Surface Aided Wireless Systems with Imperfect Hardware," *Electronics*, vol. 11, no. 6, Jan. 2022, Art. no. 900, https://doi.org/10.3390/electronics11060900.

[20] F. Fang, Y. Xu, Q. V. Pham, and Z. Ding, "Energy-Efficient Design of IRS-NOMA Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14088–14092, Aug. 2020, https://doi.org/10.1109/TVT.2020.3024005.

[21] G. Li, M. Zeng, D. Mishra, L. Hao, Z. Ma, and O. A. Dobre, "Energy-Efficient Design for IRS-Empowered Uplink MIMO-NOMA Systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9490–9500, Sep. 2022, https://doi.org/10.1109/TVT.2022.3177132.

[22] W. Du *et al.*, "Weighted Sum-Rate and Energy Efficiency Maximization for Joint ITS and IRS Assisted Multiuser MIMO Networks," *IEEE Transactions on Communications*, vol. 70, no. 11, pp. 7351–7364, Aug. 2022, https://doi.org/10.1109/TCOMM.2022.3213356.

[23] Y. Qi and M. Vaezi, "IRS-Assisted Physical Layer Security in MIMO-NOMA Networks," *IEEE Communications Letters*, vol. 27, no. 3, pp. 792–796, Mar. 2023, https://doi.org/10.1109/LCOMM.2023.3235722.

[24] T. Zhang and S. Mao, "Energy-Efficient Federated Learning With Intelligent Reflecting Surface," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, pp. 845–858, Jun. 2022, https://doi.org/10.1109/TGCN.2021.3126795.

[25] A. Ihsan, W. Chen, M. Asif, W. U. Khan, Q. Wu, and J. Li, "Energy-Efficient IRS-Aided NOMA Beamforming for 6G Wireless

Communications," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 4, pp. 1945–1956, Sep. 2022, https://doi.org/10.1109/TGCN.2022.3209617.

[26] S. H. Hong, J. Park, S. J. Kim, and J. Choi, "Hybrid Beamforming for Intelligent Reflecting Surface Aided Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7343–7357, Sep. 2022, https://doi.org/10.1109/TWC.2022.3157880.

[27] Y. Wu, F. Zhou, W. Wu, Q. Wu, R. Q. Hu, and K. K. Wong, "Spectrum and Energy Efficiency Tradeoff in IRS-Assisted CRNs with NOMA: A Multi-Objective Optimization Framework," in *ICC 2022 - IEEE International Conference on Communications*, Seoul, Korea (South), Feb. 2022, pp. 3930–3935, https://doi.org/10.1109/ICC45855.2022.9838851.

[28] L. Du, W. Zhang, J. Ma, and Y. Tang, "Reconfigurable Intelligent Surfaces for Energy Efficiency in Multicast Transmissions," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 6266–6271, Jun. 2021, https://doi.org/10.1109/TVT.2021.3080302.

[29] L. Li, D. Ma, H. Ren, P. Wang, W. Lin, and Z. Han, "Toward Energy-Efficient Multiple IRSs: Federated Learning-Based Configuration Optimization," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, pp. 755–765, Jun. 2022, https://doi.org/10.1109/TGCN.2021.3136306.

[30] Q. Liu, J. Wu, L. Hu, S. Bi, W. Ji, and R. Yang, "Optimal Energy Efficiency Used DDPG in IRS-NOMA Wireless Communications," *Symmetry*, vol. 14, no. 5, May 2022, Art. no. 1018, https://doi.org/10.3390/sym14051018.

[31] K. Li, C. Huang, Y. Gong, and G. Chen, "Double Deep Learning for Joint Phase-Shift and Beamforming Based on Cascaded Channels in RIS-Assisted MIMO Networks," *IEEE Wireless Communications Letters*, vol. 12, no. 4, pp. 659–663, Apr. 2023, https://doi.org/10.1109/LWC.2023.3238073.

[32] P. Saikia, S. Pala, K. Singh, S. K. Singh, and W. J. Huang, "Proximal Policy Optimization for RIS-assisted Full Duplex 6G-V2X Communications," *IEEE Transactions on Intelligent Vehicles*, pp. 1–16, 2023, https://doi.org/10.1109/TIV.2023.3275632.

[33] H. B. Mahesh, G. F. A. Ahammed, and S. M. Usha, "Design and Performance Analysis of Massive MIMO Modeling with Reflected Intelligent Surface to Enhance the Capacity of 6G Networks," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12068–12073, Dec. 2023, https://doi.org/10.48084/etasr.6234.