symmetry

MDPI

Essay

Optimal Energy Efficiency Used DDPG in IRS-NOMA Wireless Communications

Quanjin Liu, Jianlan Wu, Langtao Hu *, Songjiao Bi, Wen Ji and Rui Yang

School of Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, Anqing 246011, China; liuquanjin@aqnu.edu.cn (Q.L.); wujianlan@aqnu.edu.cn (J.W.); bisongjiao@aqnu.edu.cn (S.B.); jw18130432787@163.com (W.J.); yang_rui@aqnu.edu.cn (R.Y.)
* Correspondence: hulangtao@aqnu.edu.cn

Abstract: Combining Intelligent Reflecting Surface (IRS) with Non-Orthogonal Multiple Access (NOMA) technology is a viable option for increasing communication performance. Firstly, a NOMA downlink transmission system assisted by IRS is established in this study, for maximizing its energy efficiency. Then a Deep Deterministic Policy Gradient (DDPG) algorithm with symmetric properties is used to further optimize the energy efficiency of the system by intelligently adjusting the beam-forming matrix of the access point (AP) and the phase-shift matrix of the IRS. According to the simulation results, the proposed IRS-assisted NOMA downlink network based on the DDPG algorithm presented a considerably higher energy efficiency than the orthogonal multiple access network.

Keywords: intelligent reflecting surface; deep deterministic policy gradient; non-orthogonal multiple access

Citation: Liu, Q.; Wu, J.; Hu, L.; Bi, S.; Ji, W.; Yang, R. Optimal Energy Efficiency Used DDPG in IRS-NOMA Wireless Communications. *Symmetry* **2022**, *14*, 1018. https://doi.org/10.3390/sym14051018

Academic Editors: Chun-Yen Chang, Teen-Hang Meen, Charles Tijus and Po-Lei Lee

Received: 18 April 2022
Accepted: 14 May 2022
Published: 17 May 2022

1. Introduction

People's expectations for mobile data transfer rates have grown drastically with the massive expansion of Internet users and the rapid development of the Internet of Things (IoT) in recent years. According to the white paper by Cisco [1], the mobile network connection speeds will grow more than triple in 2023, reaching 43.9 Mbps. Some prospective technologies, such as Intelligent Reflecting Surface (IRS) [2–5], Non-Orthogonal Multiple Access (NOMA) [6–9] and Deep Reinforcement Learning (DRL) algorithm [10–13], have already been developed and explored to boost the user's transmission rate.

IRS, as a novel technique, has reshaped the wireless transmission environment in recent years by altering passive beamforming. It is composed of a variety of inexpensive, passively reflecting components that may be used in current wireless networks [14]. NOMA is another revolutionary technology of the 5th generation mobile communications, which utilized the superposition coding of transmitter and interference elimination of receiver to improve system throughput and spectral efficiency. This technology, unlike the classic Orthogonal Multiple Access (OMA) technologies, assigns a fraction of subcarriers to a specific user, allowing Base Stations (BS) to broadcast concurrently with multiple mobile users [15]. How to integrate IRS and NOMA effectively has now been a focus of the wireless communications improvements.

IRS is now widely used in wireless communication systems. In [16], it investigated the IRS-assisted downlink multi-input multi-output (MIMO) system. In [17], the paper studied an IRS-enhanced orthogonal frequency division multiplexing (OFDM) system. IRS-assisted UAV wireless communication systems have been widely proposed in [18,19]. The IRS-assisted NOMA wireless communication systems have also attracted extensive attention from researchers in [20–26].

Ding proposed a simple IRS-aided NOMA transmission scheme, which can provide services for users at the edge of the cell [21]. Wang investigated the effectiveness of IRS in a

NOMA system in terms of transmitting power consumption [22]. In [23], it highlighted the primary responsibilities of IRSs in MIMO-NOMA systems, in contrast to the prior research. A three-step approach to innovative resource allocation was also presented to meet this demand [24]. In [25], it explored how IRS might be used in NOMA, where a BS transmitted the superposed signals to several users via an IRS. In [26], it described an energy-efficient method that struck a fair balance between sum-rate maximization and overall power usage reduction.

Reinforcement Learning (RL) has been widely used in communications. Even in time-varying channels, DRL may tackle wireless communication challenges by exchanging information and reward mechanisms with the communication environment. DRL was used in [27] to investigate the design of integrating the beamforming matrix and the IRS phase-shift matrix. In [28], a Deep Deterministic Policy Gradient (DDPG) technique was applied to intelligently change the phase-shift matrix of AP by controlling numerous Reflection Elements (RE) of the IRS.

Accordingly, this paper proposed an energy efficiency optimization algorithm based on DRL for an IRS-assisted NOMA downlink wireless communications system. Since wireless communications were both time-varying and continuous, the DDPG algorithm from DRL was selected to specifically optimize the energy efficiency of the IRS-NOMA communications system. The main contributions of this paper are as follows:

(1) We studied the IRS-NOMA downlink wireless communication systems, considering the base station and users' direct transmission link to maximize the system energy efficiency.
(2) Based on the DDPG algorithm, the beam forming matrix of AP and phase shift matrix of IRS are jointly optimized to maximize system energy efficiency.
(3) Compared with the conventional OMA networks, the proposed algorithm presented a higher energy efficiency.

2. System Model

2.1. IRS-NOMA System Model

An IRS-assisted NOMA wireless communication (IRS-NOMA) is consisting of an IRS, a BS and multiple users, as shown in Figure 1. BS has a set of antennas, denoted by $\mathcal{M} = \{1, 2, \ldots, M\}$. $\mathcal{K} = \{1, 2, \ldots, K\}$ is to denote the set of users. The IRS consists of N-number reflection units, denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$. IRS can independently reflect the received signal through Channel State Information (CSI) and alter the amplitude or phase of the reflection unit to coordinate the signal's directions.

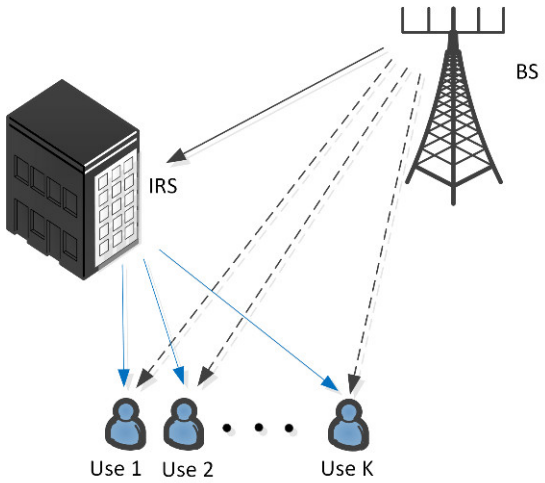

Figure 1. IRS-NOMA system model.

The IRS-NOMA wireless communication is a downlink system. Its transmitted signal can be expressed as:

$$s = \sum_{i=1}^{K} \boldsymbol{\omega}_i x_i, \tag{1}$$

where $\boldsymbol{\omega}_i \in \mathbb{C}^A$ is the generated precoding matrix, $x \in \mathbb{C}^A$ represents the signals to be sent [29].

The users' received signal is addressed by:

$$y_k = (\mathbf{H}_1 \boldsymbol{\Theta} \mathbf{G} + \mathbf{H}_2)s + e_k, \tag{2}$$

where the channel parameters from BS to the user, BS to IRS, and IRS to the user are signified as $\mathbf{H}_2 \in \mathbb{C}^{M\times 1}, \mathbf{G} \in \mathbb{C}^{N\times M}$, and $\mathbf{H}_1 \in \mathbb{C}^{N\times 1}$, respectively. IRS's diagonal phase-shift matrix is expressed as $\boldsymbol{\Theta} = diag(e^{j\theta_1}, \ldots, e^{j\theta_N}) \in \mathbb{C}^{N\times N}$, with $\theta_n \in [0, 2\pi]$, $n \in \{1, \ldots, N\}$. $e_k \sim \mathcal{CN}(0, \sigma^2)$ is the Additive White Gaussian Noise (AWGN).

Users are ranked in terms of channel quality compared with BS to simplify NOMA transmission. According to NOMA technology's decoding principle, users must first decode and exclude individuals whose channel quality is lower than their own until their signals are decoded later, while signals transmitted to users whose channel quality is higher than their own must be treated as noises. As a result, the signal that the user received can be expressed as:

$$y_k = (\mathbf{H}_1 \boldsymbol{\Theta} \mathbf{G} + \mathbf{H}_2)s' + e_k, \tag{3}$$

where $s' = \sum_{i=k}^{K} \boldsymbol{\omega}_i x_i$ is the ranked signal.

After decoding the signals successfully, the transmission rate of the user k will be specified by:

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

where $\frac{|(\mathbf{H}_1 \boldsymbol{\Theta} \mathbf{G}+\mathbf{H}_2)\boldsymbol{\omega}_k|^2}{\left|(\mathbf{H}_1 \boldsymbol{\Theta} \mathbf{G}+\mathbf{H}_2)\sum_{j=k+1}^{K} \boldsymbol{\omega}_k\right|^2+e_k^2}$ is the Signal-to-Interference-plus-Noise Ratio (SINR) at the k-th user. Thus, the system's Spectral Efficiency (SE) can be written as:

$$\mathcal{R} = \sum_{k=1}^{K} \mathcal{R}_k. \tag{5}$$

2.2. Optimization Problem

In the IRS-NOMA system, the goal to achieve is optimizing its energy efficiency, which will be calculated as the ratio of the system's SE to total power consumption as follows:

$$\begin{aligned} \max_{\Phi, G} \eta(\mathbf{G}(t), \boldsymbol{\Phi}(t), \mathbf{H}_1, \mathbf{H}_2) &= \frac{\mathcal{R}}{\mu\sum_{k=1}^{K} p_k + P_{BS} + KP_U + NP_n} \\ s.t\ \mathrm{tr}\{\mathbf{G}\mathbf{G}^{\mathcal{H}}\} &\leq \mathbf{P}_t \\ |\phi_n| &= 1\ \forall n = 1, 2, \ldots, N \end{aligned} \tag{6}$$

where $\mu\sum_{k=1}^{K} p_k, P_{BS}, P_U, P_n$ is the hardware power consumed by BS transmitting, BS power loss, mobile user terminal's power loss and IRS, respectively.

3. Deep Reinforcement Learning Algorithm Theory

3.1. DRL

RL has a strong ability for environment interaction, and the interaction process between agent and environment can be expressed by Markov Decision Processes (MDPs). During the interaction process, the agent performs the action under the guidance of the strategy π generated by the RL algorithm according to the currently observed state, gets

feedback with reward from the environment, and then enters the next state. The RL algorithm can repeat the above steps and get the cumulative rewards. The purpose of the RL algorithm is to find the best method for generating the most cumulative rewards throughout the interactions with the environment.

RL's fundamental model is made up of two parts: agent and environment, including the state, action and reward. MDPs policies are related only to the current state and can be represented as: $W = \{S, A, P, R, \gamma\}$.

(1) $S = \{s_1, s_2, \ldots, s_n\}$ represents the state sets;
(2) $A = \{a_1, a_2, \ldots, a_n\}$ represents the action sets;
(3) $P^a_{s \to s'} = P_r(s' \big| (s, a))$ represents the probability that the current state s will move to the next state S' after taking the action a;
(4) $R(s, a) = \mathbb{E}[R_{t+1} | s, a]$ represents the immediate reward generated by the agent performing action a in the current state s;
(5) γ represents the discount factor, according to which each reward is assigned with a different weight.

The cumulative reward of the agent is:

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

The agent's optimization aim is to discover a strategy $\pi(a|s)$ that maximizes the total reward $G(s)$, where $\pi(a|s) = P(a|s)$, which is the probability of taking action in the current state. The following state value function is obtained by:

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

The function is the expectation of reward based on the state S at the time t. The expected reward of state at time t after selecting action a is called the state-action value function. The following is its calculation formula:

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

The ideal equation of the following two ones may be solved through the Bellman optimal criterion:

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

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

where $V_*(s)$ is the value of selecting the optimal action considering all possible ones under the current state; $Q_*(s, a)$ is the long-term value of considering all possible states after performing an action in each state and then choosing the best one to perform in those states.

3.2. DDPG Algorithm

Since various parameters of the wireless cellular network are constantly changing and the action values of the agent in the DDPG algorithm are continuous, it is selected in this study to optimize the energy efficiency of the IRS-NOMA system. The block diagram of the IRS-NOMA downlink system's optimization model based on the DDPG algorithm is shown in Figure 2. DDPG is divided into two sections: actor and critic. The actor-network and actor target network are included in the actor section, whereas the critic network and critic target network are included in the critic section, respectively. So, the DDPG algorithm has symmetric properties. All four above are Deep Neural Networks (DNNs) which have

three layers: input layer, hidden layer, and output layer. The DNN can have generalization ability from its multiple hidden layers.

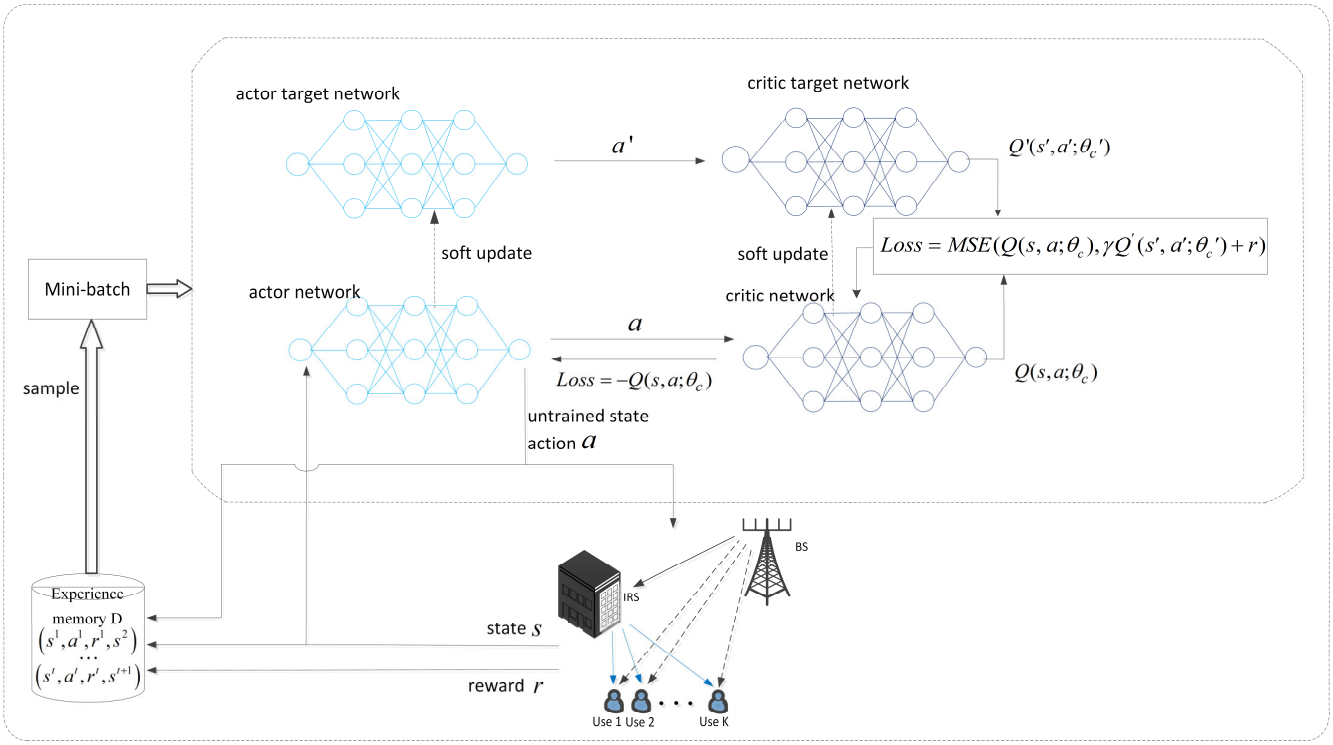

Figure 2. The block diagram of the IRS-NOMA downlink system's optimization model, based on the DDPG algorithm.

The actor-network outputs the agent action according to the input state of the agent, then the critic network outputs the Q values according to the different input states and actions of the actor-network. In the learning process, the critic network serves as an auxiliary network for reviewing the outputs of the actor-network. It is in charge of analyzing the performance of the actor-network but does not participate in generating actions.

The DDPG algorithm seeks to maximize the reviewing value of the critic network which makes an accurate evaluation of the actor-network as far as possible. The actor-network and critic-network work together to optimize the parameters of the neural network with two parts according to the following formula:

$$\theta_a^* = \underset{\theta_a}{\operatorname{argmax}} Q(s, a; \theta_c|_{a=A(s,\theta_a)}) \tag{12}$$

and

$$\theta_c^* = \underset{\theta_c}{\operatorname{argmin}} \frac{1}{2} \left(Q(s, a; \theta_c)\Big|_{a=A(s,\theta_a)} - R(s, a) \right)^2. \tag{13}$$

θ_a^* and θ_c^* are differentiable. According to the chain rule, their gradients can be obtained as follows:

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

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

The soft update is adopted for target network parameters which are slowly updated at the beginning. The formulas are as follows:

$$\theta_a' = \tau \theta_a + (1 - \tau) \theta_a', \tag{16}$$

$$\theta'_c = \tau\theta_c + (1-\tau)\theta'_c, \tag{17}$$

where τ is the update coefficient.

4. Application of DDPG in IRS-NOMA Wireless Communications

In this paper, the DDPG algorithm is adopted to optimize the energy efficiency of the IRS-NOMA wireless communication system. The state, action, and instant reward are as follows since the DDPG algorithm's agent can gather real-time information on channel status:

(1) State: The channel matrix $\mathbf{H}_1$ and $\mathbf{H}_2$, the transmission power at the tth step, the received power of users at the tth step and the action from the $(t-1)$th step jointly play a role in determining the state $s^{(t)}$ at the tth step. Since the matrices of the system are complex, information is lost by absolute operators during calculating the transmitted and received powers, so it will be necessary to distinguish the real and imaginary parts of the transmitted signal.

(2) Action: The transmit beamforming matrix $\mathbf{G}$ and the phase shift matrix $\mathbf{\Phi}$ are applied to create the action. Similarly, $\mathbf{G} = Re\{\mathbf{G}\} + Im\{\mathbf{G}\}$ and $\mathbf{\Phi} = Re\{\mathbf{\Phi}\} + Im\{\mathbf{\Phi}\}$ are also divided into real parts and imaginary parts, respectively.

(3) Reward: At the tth step of DRL, the reward will be determined as the energy efficiency $\eta(\mathbf{G}(t), \mathbf{\Phi}(t), \mathbf{H}_1, \mathbf{H}_2)$ of the IRS-NOMA wireless communications.

Algorithm complexity analysis: the DDPG algorithm has four neural networks, assuming that each one has L layers, and each layer has M_l neurons, due to their different types, the corresponding complexity will not be the same. This paper assumes that the sum nodes of BN, 'Relu' and 'tanh' layers are M_b, M_r And M_t, respectively. M_a represents the nodes of the actor-network and M_c represents the nodes of the critic network. Individual "BN" nodes, "Relu" nodes, and "tanh" nodes require 5, 1, and 6 floating-point calculations, respectively [30]. In the training process, the actor-network and the critic network will work together. So, their complexities are $\mathcal{O}(5M_b^c + M_r^c + 6M_t^c + \sum_{l=0}^{L-1} M_l^c \cdot M_{l+1}^c)$ and $\mathcal{O}(5M_b^a + M_r^a + 6M_t^a + \sum_{l=0}^{L-1} M_l^a \cdot M_{l+1}^a)$. Meanwhile, the complexity of the target network is $\mathcal{O}(\sum_{l=0}^{L-1} M_l^c \cdot M_{l+1}^c) + \mathcal{O}(\sum_{l=0}^{L-1} M_l^a \cdot M_{l+1}^a)$. Therefore, the total complexity of the algorithm will be:

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

where N, T were given as the number of training episodes and the number of steps per training round, respectively.

The pseudocode of the DDPG algorithm used in the IRS-NOMA wireless communication system is shown in Algorithm 1.

Algorithm 1: DDPG algorithm used in IRS-NOMA wireless communications system.

Input : Episode, actor − network learning rate η_a, critic network learning rate η_c, $\mathbf{H}_1$ *and* $\mathbf{H}_2$
Initialization : experience memory with size D, training actor − network parameter θ_a, target actor − network parameter $\theta_a\prime = \theta_a$, training critic network with parameter θ_c, target critic network with parameter $\theta_c\prime = \theta_c$, transmit beamforming matrix $\mathbf{G}$, phase shift matrix $\mathbf{\Phi}$
for $n = 1$ to N **do**
  Collect and preprocess $\mathbf{H}_1$ *and* $\mathbf{H}_2$ to obtain the first state s^1
  for $t = 1$ to T **do**
    Obtain action a^t from the actor-network
    Execute action a^t, observe instant reward r^t
    Observe new state s^{t+1}
  Store the experience (s^t, a^t, r^t, s^{t+1}) in the replay memory
    Calculate critic gradient $\nabla\theta_c$ by (15), and update parameter $\theta_c \leftarrow \theta_c - \eta_c \nabla\theta_c$
    Calculate actor gradient $\nabla\theta_a$ by (14), and update parameter $\theta_a \leftarrow \theta_a + \eta_a \nabla\theta_a$
  Updated actor and critic network state $s^t = s^{t+1}$
  end for
end for
Output: action, Q value function

5. Simulation Results

In this chapter, this paper simulated and quantified the energy efficiency performance of DDPG in the IRS-NOMA wireless communications system. This paper considered the Small-scale Rayleigh Fading between the BS and users. The simulation parameter settings are shown in Table 1. All the simulations are performed on a desktop with an Inter(R) Xeon(R) Platinum 8268 CPU @2.90 GHz and 16 GB memory. The simulation environment was based on tensorflow1.13. The proposed algorithm adopted the Python programming language.

Table 1. The simulation parameter settings.

Parameter	Value	Description
θ_a	0.001	the learning rate for actor-network update
θ_c	0.001	the learning rate for critic network update
D	100,000	the buffer size for experience replay
N	1000	the number of training episodes
T	10,000	the number of steps in each training episode
P_{BS}	9 dBW	the Circuit dissipated power at BS
μ	1.2	the Circuit dissipated power coefficients at BS
P_{UE}	10 dBW	the Circuit dissipated power coefficients at BS
P_n	10 dBW	the Dissipated power at the n-th IRS element

The DDPG algorithm's training diagram is shown in Figure 3. The reward value grows with the number of steps taken during the training phase and then converges into a constant amount. When the transmission power turns constant, the increasing number of users will reduce the system's energy efficiency.

As shown in Figure 4, when the transmitting power of the BS is increased, the energy efficiency will decrease accordingly. As the number of reflection units grows, the overall system's energy efficiency will fall. IRS' performance can be improved more effectively if the number of reflecting parts is increased appropriately.

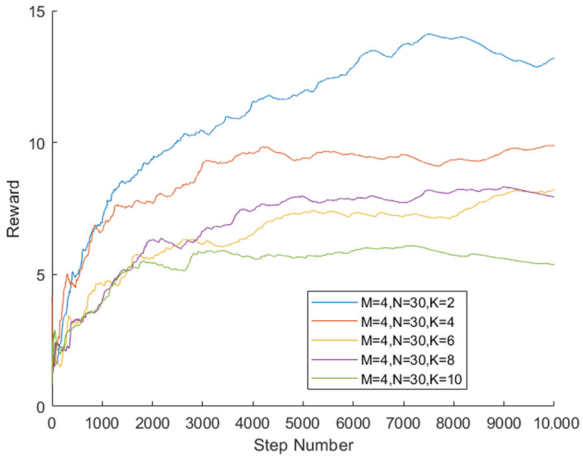

Figure 3. The DDPG algorithm's training diagram.

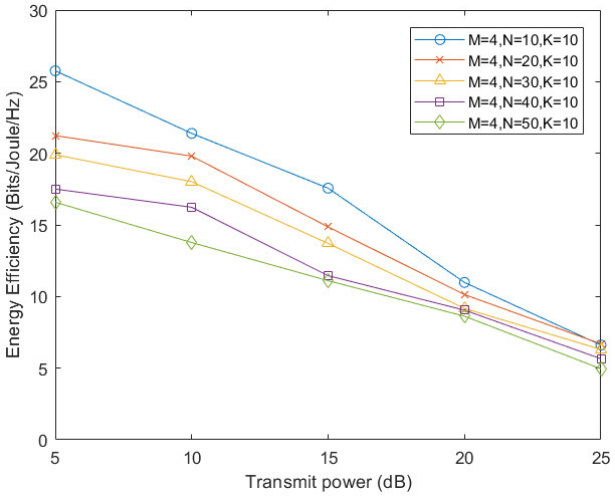

Figure 4. The diagram of energy efficiency variations with BS transmitted power fluctuations.

Figure 5 shows the relationship between transmitting power and energy efficiency among different users. The energy efficiency of the system decreases with the increase of transmitting power, as seen in the figure. The number of users also has a certain effect on the energy efficiency of the system.

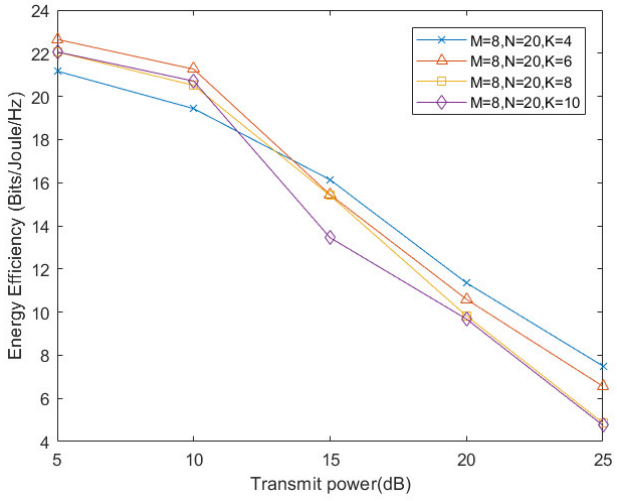

Figure 5. The diagram of energy efficiency variations with user transmitted power fluctuations.

Figure 6 shows the comparison between NOMA transmission and OMA transmission. The efficiency performance disparity between them widens as the transmitting power grows. Therefore, the addition of NOMA technology is proved to have positive significance for energy efficiency improvement.

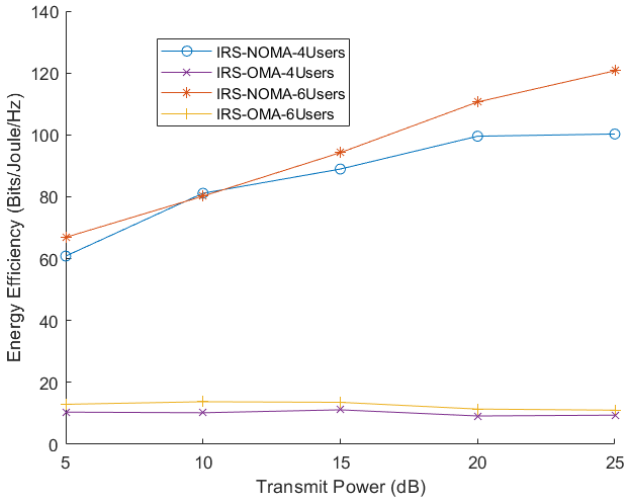

Figure 6. The influence of different transmission powers on energy efficiency.

6. Conclusions

This paper proposed an optimized IRS-NOMA communication system based on the DDPG algorithm. Aiming at the energy efficiency of this downlink wireless communication system, the DDPG algorithm from DRL was applied to optimize both the IRS's phase shifts and the beamforming vectors for finally improving the energy efficiency of the entire system. The proposed algorithm's applicability was demonstrated by the simulation results that the integration of NOMA and IRS can effectively improve the energy efficiency of IRS-NOMA wireless communication systems.

Author Contributions: Conceptualization, Q.L. and L.H.; methodology, Q.L.; software, J.W.; validation, Q.L., J.W. and L.H.; formal analysis, Q.L.; investigation, L.H.; resources, L.H.; data curation, J.W. and W.J.; writing—original draft preparation, J.W.; writing—review and editing, Q.L., J.W. and L.H.; visualization, S.B.; supervision, W.J.; project administration, R.Y.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 62171002, and the Natural Science Foundation Project of Anhui Province, grant number KJ2019A0554.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Cisco. Cisco. Cisco Annual Internet Report (2018–2023). In *White Paper*; Cisco: San Jose, CA, USA, 2020.
2. Ding, Z.; Fan, P.; Poor, H.V. Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. *IEEE Trans. Veh. Technol.* **2016**, *65*, 6010–6023. [CrossRef]
3. Liu, Y.; Qin, Z.; Elkashlan, M.; Ding, Z.; Nallanathan, A.; Hanzo, L. Nonorthogonal multiple access for 5G and beyond. *Proc. IEEE* **2017**, *105*, 2347–2381. [CrossRef]
4. Islam, S.R.; Avazov, N.; Dobre, O.A.; Kwak, K.S. Power-domain nonorthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 721–742. [CrossRef]
5. Mounchili, S.; Hamouda, S. Pairing distance resolution and power control for massive connectivity improvement in NOMA systems. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4093–4103. [CrossRef]

6. Sun, Z.; Jing, Y. On the performance of multi-antenna IRS-assisted NOMA networks with continuous and discrete IRS phase shifting. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 3012–3023. [CrossRef]
7. Rehman, H.U.; Bellili, F.; Mezghani, A.; Hossain, E. Joint active and passive beamforming design for IRS-assisted multi-user MIMO systems: A VAMP-based approach. *IEEE Trans. Commun.* **2021**, *69*, 6734–6749. [CrossRef]
8. Gong, S.; Lu, X.; Hoang, D.T.; Niyato, D.; Shu, L.; Kim, D.I.; Liang, Y.C. Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2283–2314. [CrossRef]
9. Wu, Q.; Zhang, R. Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 5394–5409. [CrossRef]
10. Feng, K.; Wang, Q.; Li, X.; Wen, C.K. Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 745–749. [CrossRef]
11. Liu, J.; Ahmed, M.; Mirza, M.A.; Khan, W.U.; Xu, D.; Li, J.; Aziz, A.; Han, Z. RL/DRL meets vehicular task offloading using edge and vehicular cloudlet: A survey. *IEEE Internet Things J.* **2022**. [CrossRef]
12. Wang, X.; Zhang, Y.; Shen, R.; Xu, Y.; Zheng, F.C. DRL-Based energy-efficient resource allocation frameworks for uplink NOMA systems. *IEEE Internet Things J.* **2020**, *7*, 7279–7294. [CrossRef]
13. Shi, J.; Du, J.; Shen, Y.; Wang, J.; Yuan, J.; Han, Z. DRL-Based V2V computation offloading for blockchain-enabled vehicular networks. *IEEE Trans. Mob. Comput.* **2022**. [CrossRef]
14. Williams, R.J.; Carvalho, E.D.; Marzetta, T.L. A communication model for large intelligent surface. *arXiv* **2019**, arXiv:1912.06644.
15. Xiao, C.; Zeng, J.; Ni, W.; Su, X.; Liu, R.P.; Lv, T.; Wang, J. Downlink MIMO-NOMA for ultra-reliable low-latency communications. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 780–794. [CrossRef]
16. Ning, B.; Chen, Z.; Chen, W.; Fang, J. Beamforming optimization for intelligent reflecting surface assisted MIMO: A sum-path-gain maximization approach. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1105–1109. [CrossRef]
17. Yang, Y.; Zheng, B.; Zhang, S.; Zhang, R. Intelligent reflecting surface meets OFDM: Protocol design and rate maximization. *IEEE Trans. Commun.* **2020**, *68*, 4522–4535. [CrossRef]
18. Yu, J.; Liu, X.; Gao, Y.; Zhang, C.; Zhang, W. Deep learning for channel tracking in IRS-assisted UAV communication systems. *IEEE Trans. Wirel. Commun.* **2022**. [CrossRef]
19. Wang, D.; Zhao, Y.; He, Y.; Tang, X.; Li, L.; Zhang, R.; Zhai, D. Passive beamforming and trajectory optimization for reconfigurable intelligent surface-assisted UAV secure communication. *Remote Sens.* **2021**, *13*, 4286. [CrossRef]
20. Yang, G.; Xu, X.; Liang, Y. Intelligent reflecting surface assisted non-orthogonal multiple access. *arXiv* **2019**, arXiv:1907.03133.
21. Ding, Z.; Poor, H.V. A simple design of IRS-NOMA transmission. *IEEE Commun. Lett.* **2020**, *24*, 1119–1123. [CrossRef]
22. Wang, H.; Liu, C.; Shi, Z.; Fu, Y.; Song, R. On power minimization for IRS-Aided downlink NOMA systems. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1808–1811. [CrossRef]
23. De Sena, A.S.; Carrillo, D.; Fang, F.; Nardelli, P.H.J.; Da Costa, D.B.; Dias, U.S.; Ding, Z.; Papadias, C.B.; Saad, W. What role do intelligent reflecting surfaces play in multi-antenna non-orthogonal multiple access? *IEEE Wirel. Commun.* **2020**, *27*, 24–31. [CrossRef]
24. Zuo, J.; Liu, Y.; Qin, Z.; Al-Dhahir, N. Resource allocation in intelligent reflecting surface assisted NOMA systems. *IEEE Trans. Commun.* **2020**, *68*, 7170–7183. [CrossRef]
25. Yue, X.; Liu, Y. Performance analysis of intelligent reflecting surface assisted NOMA networks. *IEEE Trans. Wirel. Commun.* **2022**. [CrossRef]
26. Fang, F.; Xu, Y.; Pham, Q.V.; Ding, Z. Energy-efficient design of IRS-NOMA networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 14088–14092. [CrossRef]
27. Huang, C.; Mo, R.; Yuen, C. Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 1839–1850. [CrossRef]
28. Yang, Z.; Liu, Y.; Chen, Y.; Zhou, J.T. Deep Reinforcement Learning for RIS-Aided Non-Orthogonal Multiple Access Downlink Networks. In Proceedings of the GLOBECOM 2020–2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [CrossRef]
29. Yang, Z.; Liu, Y.; Chen, Y.; Al-Dhahir, N. Machine learning for user partitioning and phase shifters design in RIS-aided NOMA Networks. *IEEE Trans. Commun.* **2021**, *69*, 7414–7428. [CrossRef]
30. Zhong, R.; Liu, Y.; Mu, X.; Chen, Y.; Song, L. AI empowered RIS-assisted NOMA networks: Deep learning or reinforcement learning? *IEEE J. Sel. Areas Commun.* **2022**, *40*, 182–196. [CrossRef]