



Deep Reinforcement Learning-Based Adaptive Discrete Phase-Shift Optimization for Intelligent Reflecting Surface Supported Wireless Communications

Hamood Shehab Hamid¹, Muthna J. Fadhil^{1*}, F. Richard Yu², Haider W. Olewi³

¹ Middle Technical University, Electrical Engineering Technical College, Baghdad 10011, Iraq

² School of Information Technology, Ottawa K1S 5B6, Canada

³ Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, UK

Corresponding Author Email: muthanafadhil@mtu.edu.iq

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.590915>

ABSTRACT

Received: 20 July 2026

Revised: 17 September 2026

Accepted: 26 September 2026

Available online: 30 September 2026

Keywords:

intelligent reflecting surfaces, deep Q-network, deep reinforcement learning, spectral efficiency enhancement, wireless channel optimization

Intelligent reflecting surfaces (IRS) have emerged as an innovative technology that has the potential to alter wireless propagation parameters and greatly enhance performance in modern communication systems. However, determining the optimal discrete phase-shift design for large IRS panels is a challenging high-dimensional optimization problem, particularly in situations with minimal feedback and dynamic channels. This study proposes a deep reinforcement learning (DRL) framework based on a deep Q-network (DQN) to tackle this challenge. This design allows for the learning of the best IRS phase-shift technique in a model-free way without needing to have a conventional optimization procedure or an explicit channel model. The DQN agent interacts with the wireless environment while iteratively changing its strategy to maximize end-to-end connection quality. The simulation results show that the proposed DQN-assisted IRS setup works effectively, outperforming the usual optimization methods. It seems to do better overall, even compared to the more traditional way. Specifically, the technique outperforms Random and Fixed- phase methods by achieving a received SNR of around 17 dB and a spectrum efficiency of roughly 4.73 bit/s/Hz. Furthermore, the learning process exhibits continuous convergence after around 120 episodes, proving the robustness of the proposed method. Overall, the results show how effectively DRL and IRS technologies can be used to create wireless communication environments that are flexible, scalable, and high-performance.

1. INTRODUCTION

The rapid expansion of wireless communication systems and the increasing need for higher data rates, more coverage, and improved energy efficiency have sparked extensive study into technologies that may alter the physical propagation environment. Among these developing technologies, intelligent reflecting surfaces (IRS) have gained much attention as an interesting framework for programmable electromagnetic-wave manipulation using low-cost, passive reflecting materials [1-3]. IRSs may provide increased spectrum efficiency, minimize interference, and enable constructive propagation of signals by adaptively adjusting the phase offsets of the passive reflecting components. Thus, IRSs have already emerged as a rather promising technology to fulfill the performance requirements of the future 6G wireless networks [4, 5]. The best IRS phase-shift design, however, turns out to be a difficult optimization problem that is non-convex and practical, having discrete-phase constraints and time varying wireless channels [6-8]. The semidefinite relaxation, the alternating optimization and the greedy selection approaches [9-11] are standard optimization techniques that might not result in a satisfactory local solution or process at a high cost.

These restrictions have been attempted to be minimised by several methods including, but not limited to, meta heuristic approaches like genetic algorithms (GA) and particle swarm optimisation (PSO) [12-14]. However, they generally require many iterations, have slow convergence and are poor scaling for increasing IRS numbers. Since they can learn environment-aware control policies without using explicit channel models, machine learning and deep reinforcement learning (DRL) have become effective techniques for IRS optimization in more recent times [15-18]. Some interesting results have been obtained in adaptive IRS control using a DRL-based approach, such as DRL-deep Q-network (DQN) [19-21], proximal policy optimization (PPO), and deep deterministic policy gradient (DDPG). Most studies so far, however, have been based on continuous-phase IRS models, on perfect channel state information (CSI), or require slow training convergence in high dimensional action spaces.

Furthermore, most recent designs rely on simpler IRS architectures, resulting in significant performance gap for discrete-phase, scalable and hardware-limited IRS designs [22-25]. These drawbacks demand an optimization method that is both fast and flexible, but model independent (not requiring any particular model of the IRS array), which is able to deal with discrete phase constraints, and which is able to

function under realistic channel conditions. The last few years have seen reinforcement learning for IRS/RIS optimization to be very successful [26-28]. But there are some features, such as scalability, and then the discrete phase control, that appear to be a bit tough. In this research, an optimization framework for the DQN algorithm is introduced to try to tackle those challenges gradually.

The challenges mentioned above inspire the concept of this study, which introduces a novel discrete-phase IRS configuration optimization paradigm based on DRL. The proposed method will adaptively learn the optimum IRS phase-shift strategy in an interactive manner with the wireless environment while avoiding the need for any particular CSI or closed-form channel model. The principal contributions of this work are:

- 1) A model-free DQN style framework for IRS optimization is proposed, it can learn the best discrete phase-shift policy direct from doing interactions with the wireless without need for any explicit channel model or iterative optimization routines.
- 2) A scalable discrete-phase IRS model is developed, explicitly accounting for practical hardware constraints, making the proposed approach suitable for real-world IRS implementations.
- 3) A comprehensive performance evaluation is conducted, benchmarking the proposed method against Random IRS configuration, Fixed-phase IRS, Greedy selection, PSO, GA, and Traditional Beamforming techniques.
- 4) Significant performance improvements are demonstrated, with the proposed DQN- based approach outperforming existing optimization techniques in terms of spectral efficiency, enhanced SNR scaling with IRS size, and faster convergence.
- 5) Analytical and mathematical interpretations are provided to support and explain the observed empirical performance improvements.

The rest of this paper is organized as follows. The related works are given in Section 2. In Section 3, the system model is discussed. The proposed framework for DRL is presented in Section 4. A discussion of the results of the simulations is given in Section 5. Finally, we conclude this work in Section 6 by future works.

2. RELATED WORKS

The potential of recent developments in IRS to control wireless propagation conditions using inexpensive, passive, and programmable reflecting elements has garnered a great deal of study interest [1-3]. Theoretical modeling and channel characterisation were the main focus of early IRS research, which examined how passive beamforming might boost received signal intensity, reduce interference, and increase spectral efficiency in single- user and multi-user settings [4-7]. These studies demonstrated the basic advantages of IRS-assisted systems, but they were predicated on perfect channel information and ideal assumptions. To develop IRS phase-shift configurations, a number of optimization-based methods have been put forth.

The earliest methods used to control IRS were alternating optimization and semidefinite relaxation methods [8-10]. These methods do not work well for setups because they are

expensive to compute and have non-convexity problems even though they work fine for small IRS arrays. Wireless communications assisted by the IRS also explored in other studies recently. In order to address the channel estimation issue, [11] proposed a channel estimation framework based on deep learning for an IRS-assisted wireless system. Zhou et al. [12] proposed a powerful and effective design of the transmitter in IRS-assisted wireless communication with imperfect cascaded channels. Moreover, He and Yuan [13] presented a cascaded channel estimation scheme, underscoring the significance of channel estimation for the successful deployment of IRS. However, IRS control methods like these often end up at optima and need to be recomputed over and over when the channel situation changes. People have used heuristic algorithms like PSO, GA and virtual annealing to optimize the IRS phase [14-16]. These methods are good at exploring and avoiding optima.

The problem is that their performance depends on how the parameters reset and when the IRS gets bigger it needs more iterations. Also because heuristic techniques work alone for each channel realization without learning from past experiences they are not good for real-time changes. Recently people have been looking into reinforcement learning as a way to make IRS optimization adaptive and model-free. IRS systems with phases have been shown to be more resilient when Q-learning and DQN are used [17-20]. In studies, actor-critic techniques like DDPG and PPO have been used for IRS situations with continuous phases [21, 22]. Even though other studies like [17, 25] have used DQN in IRS-assisted communications, this work is different in a few important ways. First this work focuses on the real-world behavior of discrete-phase IRS, with low-resolution phase quantization. It is not just a perfect case. The IRS control methods used here are practical. Take into account the limitations of real-world IRS systems. Second, the framework is built to stay scalable when the number of reflecting elements keeps growing, rather than collapsing as things get larger. Third, the DQN agent that we put forward learns an adaptive phase-selection rule in a model freeway, and it does that without depending on those iterative optimization methods. In short, these choices make the proposed method fit for real large-scale IRS setups, while still keeping the computational burden relatively low.

Despite these benefits, a lot of current RL-based research takes into account ideal CSI assumptions, constrained action spaces, or simplified system models. Furthermore, because of the large dimensionality of the action space, the majority of DRL-based methods suffer from poor training convergence and concentrate on very modest IRS sizes. Based on the above literature review, three key limitations can be identified: the high computational complexity of classical AO/SDR-base methods, the scalability and convergence limitations of meta-heuristic algorithms, and the limited practicality of existing DRL-based IRS frameworks under realistic hardware constraints and imperfect CSI. In contrast to prior works, this paper proposes a scalable, discrete-phase, and model-free DQN-based IRS optimization framework tailored for practical IRS hardware implementations.

The proposed approach enables the learning of an optimal phase-shift policy directly through interaction with the wireless environment, supports large IRS arrays with reduced computational complexity, and achieves superior SNR and spectral efficiency with faster and more stable convergence. These characteristics position the proposed framework as a practical and effective solution for learning-driven IRS

optimization in next-generation wireless communication systems. Table 1 shows the main differences between the proposed approach and other DRL-based IRS optimization methods in terms of learning method, phase-shift control, CSI

requirements, IRS configuration and key limitations, to further illustrate the novelty of the proposed approach relative to existing IRS optimization methods.

Table 1. Comparison with other existing IRS optimization methods based on DRL

Ref.	Method	Discrete Phase	CSI Requirement	IRS Size	Main Limitation
[29]	Optimization DRL-based IRS	Yes	CSI-based	Not emphasized	Mainly focuses on maximization MISO SNR
[30]	DQN/DDPG	Yes (DQN); continuous (DDPG)	Channel/information environment	Multiple IRSs	Increases complexity due to optimization IRS and trajectory UAV Joint
[31]	Model with free DRL	Phase adaptation	No sub-channel explicit CSI	Not emphasized	Mainly designed for configuration smart-radio
[32]	DRL-based optimization phase	Continuous	Limited CSI	Configurable IRS	Optimization of sum-rate and NOMA downlink focused
[33]	Actor-critic Deep	Continuous phase control	Correlation channel/state information	Not emphasized	Assumes environments correlated spatiotemporally
[34]	PPO-based DRL	Phase-matrix optimization	Channel/network state	Configurable IRS	Scenario D2D optimization phase and power joint
[35]	DDPG	Continuous	Channel-dependent	MIMO IRS system	Optimization phase by continuous action
Proposed DRL	DQN	Yes (2-bit)	Model-free; not CSI-free	As specified in this work	IRS size grows with discrete action space

3. METHODOLOGY

This section presents the proposed method, comprising of a system model and the framework for optimization of phase-shift by employing a DQN. It additionally offers the parameters utilized for simulation within this study.

3.1 System model

The IRS-assisted downlink wireless communication system examined in this work is described in this section. Three primary parts make up the system: A single-antenna user equipment (UE), an IRS made up of reflecting passive N pieces, and antenna base station (BS). By cleverly regulating the phase shifts of its reflecting components, the IRS is positioned between the BS and the UE to increase the received signal intensity at the user. A baseband signal x with unit average power is sent by the BS, i.e., $E[|x|^2] = 1$. Three propagation links the BS-IRS link, the IRS-UE link, and the direct BS-UE connection affect the signal received at the UE. These channels are indicated by the symbols $h_{BR} \in \mathbb{C}N \times 1$, $h_{RU} \in \mathbb{C}N \times 1$, and $h_D \in \mathbb{C}$, respectively. Figure 1 illustrates the architecture of the suggested IRS-assisted wireless system and shows the general structure of the BS-IRS-UE communication circuit.

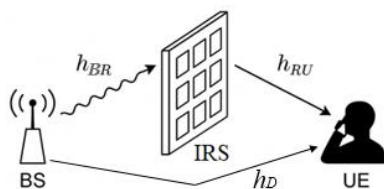


Figure 1. IRS-assisted wireless communication network system model

While the IRS receives the incident signal via the channel

h_{BR} , applies a programmed phase-shift matrix, and reflects the signal toward the UE via the channel h_{RU} , the BS broadcasts the signal toward the UE. Depending on the propagation environment, there may also be a direct BS-UE route, denoted by h_D . Without the use of active radio-frequency chains, the IRS functions as a largely passive met surface that modifies the wireless propagation environment. Constructive signal combining at the receiver is made possible by the independent phase adjustment of the incident electromagnetic wave by each IRS element. The effective received signal at the UE may be greatly increased by correctly constructing the IRS phase-shift vector, which will improve coverage performance, spectral efficiency, and SNR. Determining the ideal IRS phase-shift design that improves system efficiency under dynamic and time-varying channel circumstances is therefore the goal of the suggested optimization approach. It should be assert that the IRS is like a passive reflecting surface, and it does not really do channel estimation or baseband processing.

3.1.1 IRS architecture and unit cell description

The IRS is represented as a planar reflecting array made up of N sub-wavelength components, each of which can cause the incident electromagnetic wave to undergo a controlled phase shift. Through passive beamforming, the IRS allows for intelligent control of the wireless propagation environment by suitably adjusting the phase response of these components. A tiny metallic patch that reflects the incident electromagnetic wave usually makes up each IRS element, also called a unit cell. In order to provide discrete phase switching, a PIN diode is often placed between the patch and the control line. This allows the reflection phase to be changed between specified states based on the applied biasing condition. A substrate layer and a ground plane are used to provide appropriate electromagnetic isolation and reliable operation, while the control line provides the necessary DC bias from the IRS controller. A simplified IRS unit cell is schematically depicted in Figure 2, which captures the essential structural components required for system-level optimization.

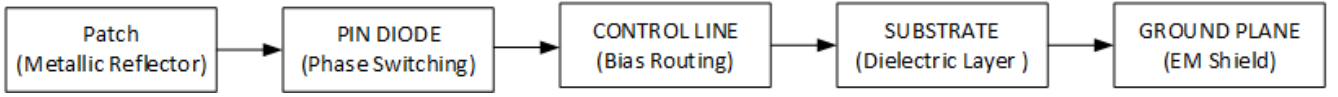


Figure 2. Simplified schematic illustration of an IRS unit cell

The reflective amplitude is simply assumed to be unity for all the IRS elements, just to make the system-level analysis a little simpler, and principally to allow focusing on the degree of accuracy; the suggested DQN-based phase shift optimization procedure is successful. In real life though, practical IRS unit cells made of PIN diodes exhibit insertion losses, amplitude fluctuations, and even reflection coefficients that depend on the phase. But in this study, those hardware quirks are mostly ignored. The DQN state is the state observed at the BS at the current time interval, which is used to determine the propagation condition. The action is one of the discrete configurations of the IRS phase-shift defined in the pre-defined phase codebook. After selecting an action, it is correlated with the corresponding phase-shift vector and then applied to the IRS elements. The definitions of state and action are summarized in Table 2.

Table 2. The IRS configuration used in the simulated environment

Item	Definition
state (st)	Current channel observations at the BS
Action (at)	Discrete IRS phase-shift configuration number selected by the DQN
Phase set	$\{0, \Delta\theta, 2\Delta\theta, \dots, (Q-1)\Delta\theta\}$
Action mapping	IRS phase-shift vector controlled by action selection.
Output	Vector phase-shift function to IRS elements.

3.1.2 Channel and IRS phase-shift model

Consider a channel model fading by quasi-static block, where the BS-IRS, IRS-UE, and direct BS-UE channels remain constant within one coherence interval and vary independently across intervals. The received baseband signal at the UE is expressed as [1, 2]

$$y = (h_{RU}^T \Theta h_{BR} + h_D) x + k \quad (1)$$

The transmitted symbol x satisfies $E[|x|^2] = 1$, and $k \sim \text{CN}(0, \sigma^2)$ is additive white Gaussian noise with variance σ^2 . The IRS phase-shift matrix is defined as

$$\Theta = \text{diag} \{e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_N}\} \quad (2)$$

where, θ_N is the phase shift applied by the N_{th} IRS reflecting element. The cascaded IRS-assisted channel is given by $h_{RU}^T \Theta h_{BR}$, and the effective BS-UE channel is expressed as

$$h_{\text{eff}} = h_{RU}^T \Theta h_{BR} + h_D \quad (3)$$

Given a transmit power P , the instantaneous received SNR at the UE is

$$\text{SNR} = \frac{P |h_{\text{eff}}|^2}{\sigma^2} \quad (4)$$

And the achievable rate (spectral efficiency) is computed as

$$R = \log_2(1 + \text{SNR}) \quad (5)$$

This serves as the proposed DQN agent's training reward function and the primary optimization objective. In practical IRS implementations, each reflecting element applies a discrete phase shift selected from a finite set

$$\Theta_m = \{0, \Delta\theta, 2\Delta\theta, \dots, (Q-1)\Delta\theta\}, m = 1, \dots, N \quad (6)$$

where, $Q = 2^b$ denotes the number of quantization levels for a b -bit phase controller and $\Delta\theta = \frac{2\pi}{Q}$. Consequently, the search space grows exponentially as Q^N , rendering the optimization problem highly non-convex and NP-hard even for moderate IRS sizes. The conventional optimization techniques, including alternating optimization, semidefinite relaxation and meta-heuristic algorithms, are subject to excessive complexity and limited adaptability due to the discrete phase constraint, high dimensionality and sensitivity to imperfect channel state information.

In practical IRS hardware designs can employ extra components, tunable and adjustable reflection using capacitors or varactors. In this work, the simplified unit cell model used is found to be a valid representation of the actual physical situation, as exposed through the use of amplitudes. essential reflection behavior needed for evaluation of performance and algorithm design. Table 3 listed the IRS configuration used in the simulated scenario, which includes some important characteristics of the IRS, such as the phase quantization resolution, Assumptions, spatial arrangement and the number of reflecting elements in the channel models. These settings were selected as they were found to be a good choice for evaluating the scalability and robustness of the proposed DQN. Optimize it under the specific hardware constraints and to simulate genuine deployment situations.

Table 3. The IRS configuration used in the simulated environment

IRS Parameter	Value
IRS type	Planar Reflective Intelligent Surface
Total number of IRS elements (N)	20, 40, 60, 80, 100, 120
Element spacing	$\lambda/2$
Phase resolution	2-bit
Phase shift quantization levels (Q)	$\{0, \pi/2, \pi, 3\pi/2\}$
Reflective amplitude	1 (ideal)
IRS deployment height	4 m
BS → IRS distance	20 m
IRS → UE distance	30 m
Channel model	Rician fading
Rician K-factor	3 dB
Path loss exponent	2.2
IRS control	DRL-based discrete phase optimization

It is important to note that the discrete IRS phase shift optimization issue may have a very wide action space and the problem becomes much worse with the increasing number of reflecting parts. For an IRS with (N) elements and (L) discrete

phase levels, the total number of potential phase configurations explodes like (L^N) . Thus an exhaustive search becomes impracticable relatively rapidly once things grow large-scale for deployments. This is somewhat addressed in the proposed DQN technique by the use of a limited phase shift codebook where each action corresponds to a possible IRS beamforming configuration, instead of a complete brute force combination phases of each single element. Moreover, the method uses experience replay along with target network updates, which helps with learning stability and also convergence. Even though the discretization we adopted is a practical compromise between complexity and spectral efficiency gains, there's still room to go further so future work will look into more scalable DRL designs such as Double DQN, Dueling DQN, or even actor-critic style methods. These challenges motivate the adoption of a learning-based optimization framework, which is presented in the following section.

This current study looks at a kind of single-user, single-antenna IRS-assisted link scenario, basically to see how well the proposed DQN based optimization framework works, in practice. If required to push the idea further, moving toward multi-user, multi-antenna setups, and also larger scale IRS deployments seems like a key next step for future research. Also, the framework should really be read as a model-free optimization method rather than something that is strictly CSI-free, because channel estimates are still used as state observations. At the same time, there is no need for an explicit channel model, and no traditional optimization routine is required.

3.2 Proposed DRL framework

In this section, DQN proposed based framework designed to optimize the IRS phase-shift vector in real time. The overall goal is to learn a mapping between the wireless environment states and the optimal IRS configuration that maximizes the received signal strength or spectral efficiency at the UE. This section introduces the proposed DQN-based IRS optimization framework, including system modeling, state/action/reward formulation, neural architecture, and training pipeline.

3.2.1 System overview

The overall architecture of the proposed DRL-based IRS optimization framework is illustrated in Figure 3. In this system, a DQN agent is employed to autonomously control the discrete phase-shift configuration of the IRS in order to enhance the wireless link performance. The agent interacts iteratively with the wireless environment by observing the system state, selecting appropriate IRS phase-shift actions, and receiving performance- based rewards that reflect the achieved link quality at the UE. In a closed-loop interaction model, the DQN agent constantly improves its policy for making decisions depending on input from the environment. The agent applies this configuration on the reflecting surface after observing the current propagation circumstances and determining a discrete phase-shift vector for the IRS elements at each time step. A reward is created to direct the learning process once the incoming signal behavior is assessed. The agent is incentivized by this payment to choose IRS settings that optimize important performance metrics like spectral efficiency and SNR.

An extensive depiction of the environment-agent interaction is shown in Figure 3. Specifically, the agent is rewarded (rt)

based on the link performance achieved after observing at time step (t) and selecting an action (at) the system state (st) that corresponds to a discrete IRS phase-shift vector. The interaction loop is completed when the environment feeds the appropriate reward back to the agent after assessing the applied action by analyzing the quality of the received signal. The suggested DRL methodology allows the wireless system to develop an effective IRS control strategy without depending on explicit CSI estimate through repeated interactions over several episodes. This learning-driven method achieves efficient and adaptable IRS phase optimization while drastically lowering computing complexity.

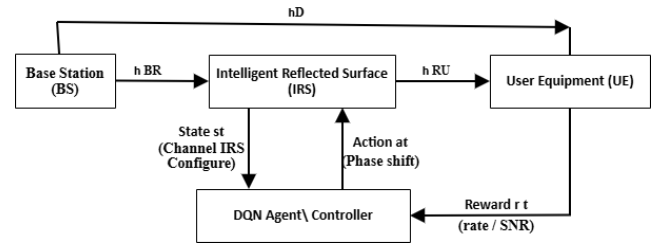


Figure 3. IRS phase-shift optimization using the suggested DRL framework based on DQN

The proposed IRS controller, as shown in Figure 3, is a closed-loop structure that observes, decides on, and actuates a closed-loop IRS. As seen from Figure 3, the proposed DQN based IRS controller works in a closed-loop observation-decision-actuation-feedback structure for a closed loop IRS. First, the controller reads one of the current state of the system, including the channel information received from the IRS. From this observation, the DQN agent chooses an appropriate action of phase shift distortion in the discrete action space. The selected action is then applied to the IRS elements to alter the reflected signal towards the UE. Transmission is followed by an estimation of the resulting system performance based on achievable rate/SNR, thus providing a reward back for the DQN agent. The feedback is then fed into a processing routine which adjusts future phase-shift actions and thus completes the closed-loop control cycle. The selected phase-shift architecture assumed in the considered architecture is that the DQN controller is placed at the BS-side, where system information is available, and the selected phase-shift architecture is sent to the IRS controller.

3.2.2 State, action, and reward design

State representation, action generation, reward calculation, and network training and updating make up the four primary parts of the technique. The agent and the wireless environment engage in several interaction episodes to carry out the entire learning process. The important observations for the IRS optimization job are captured by the system state represented by st, which is defined as follows: $s_t = \{h_{BR}, h_{RU}, h_D, \Theta_{t-1}\}$. The IRS phase-shift configuration used in the preceding time step is represented by Θ_{t-1} . The immediate propagation environment that the wireless connection experiences is characterized by these characteristics taken together. The process at is equivalent to choosing a fresh discrete phase-shift vector. The agent outputs the phase configuration at $\in \Theta^N$, where N is the number of reflecting elements, with a Q-bit phase resolution.

Even though the theoretical IRS configuration space scales in an exponential way with QN, the DQN agent that proposed

doesn't really do full, exhaustive scanning of every possible phase shift combination. Rather than brute force trying everything, it learns a phase-selection policy via ongoing back-and-forth interactions with the wireless environment, plus reward-guided updates that steer learning. That way, it can pinpoint favorable IRS configurations in a much more efficient manner, without needing to explicitly list or enumerate the entire action space. The incentive is a reflection of how the chosen IRS configuration has improved communication quality. The agent interacts with the environment at each time step t by choosing an IRS phase-shift action that influences spectral efficiency and SNR. The reward r_t is illustrated as:

$$r_t = \log_2(1 + \text{SNR}_t) \quad (7)$$

The DQN agent monitors the current system state s_t , creates an action a_t , and applies the appropriate phase-shift configuration to the IRS at each interaction step. The learning loop is completed and progressive policy improvement is made possible by measuring the resultant received signal at the UE and feeding the computed reward r_t back to the agent. The reward function is basically defined using spectral efficiency, to boost the communication performance of the IRS assisted link. The IRS controller and the biasing circuitry is assumed relatively small and nearly steady compared with the transmit power, so it is not written out in the reward formulation, explicitly though.

The channel facts needed for operation are gathered at the base station using pilot assisted estimation, and then they are relayed to the DQN agent via the IRS controller. That controller figures out the phase-shift configuration, more or less choosing the coefficients to be used. So the IRS just applies those selected terms, while all the signal processing stays on the active transceiver side.

3.2.3 DQN architecture and training procedure

Table 4. DQN hyper parameters employed for training the IRS control agent

Hyper Parameter	Value
Learning rate (α)	0.001
Discount factor (γ)	0.99
Exploration strategy	ϵ -greedy
Initial ϵ	1.0
Minimum ϵ	0.01
Decay rate ϵ	0.995
Replay buffer size	50,000
Mini-batch size	64
Target interval for network updates	Every 100 steps
Optimizer	Adam
Loss function	Mean Squared Error (MSE)
Neural network architecture	Fully-connected (128-128)
Activation function	ReLU
Training episodes	200
Steps per episode	50
Weight initialization	Glorot (Xavier) initialization
Software	MATLAB R2024a
Random seed	42
Convergence criterion	Stable reward variation ($< 1\%$) over 20 consecutive episodes

The complete workflow is explained at two complementary

levels in order to make the logical structure of the suggested RL framework clear. When taken as a whole, these procedures demonstrate how the agent and wireless environment interact iteratively and gradually converge on an efficient IRS phase-shift configuration.

Table 4 summarizes the main hyper parameters that were utilized to train the suggested DQN-based IRS optimization framework. These parameters control Q-network stability, learning dynamics, and experience replay efficiency periodic target network updates and experience replay are used with the target network parameters being updated as $\omega \rightarrow \omega'$ every predetermined number of steps.

Two neural networks make up the DQN architecture used in this framework: a target Q-network $Q'(s, a; \omega')$ and an online Q-network $Q(s, a; \omega)$. While the target network offers consistent target values throughout training, the online network is in charge of action selection. The Bellman optimality equation is used to calculate the goal value for each transition (s_t, a_t, r_t , and s_{t+1}) sampled from the replay buffer as

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a', \omega') \quad (8)$$

where, s_t and s_{t+1} refer to the system's present and subsequent states, respectively. a_t is the chosen IRS phase-shift action, r_t is the reward that was earned, a' is a candidate action in the maximization, γ is the discount factor ($0 < \gamma \leq 1$), while ω and ω' are the parameters of the online and target networks. The Bellman optimality rule is used to update the online Q-network parameters as expressed in Eq. (9):

$$Q(s_t, a_t, \omega) \rightarrow Q(s_t, a_t, \omega) + \alpha[r_t + \gamma \max_{a'} Q(s_{t+1}, a', \omega') - Q(s_t, a_t, \omega)] \quad (9)$$

where, α indicates the learning rate ($0 < \alpha \leq 1$), which regulates how much fresh data is added to the Q-value update. An ϵ -greedy strategy for exploitation and exploration strike balance, in which the agent chooses either probability ϵ with random action or probability $1 - \epsilon$ with maximizes the expected Q-value action.

$$a_t = \begin{cases} \text{random action, with probability } \epsilon \\ \arg \max_a Q(s_t, a_t, \omega), \text{ with probability } 1 - \epsilon \end{cases}$$

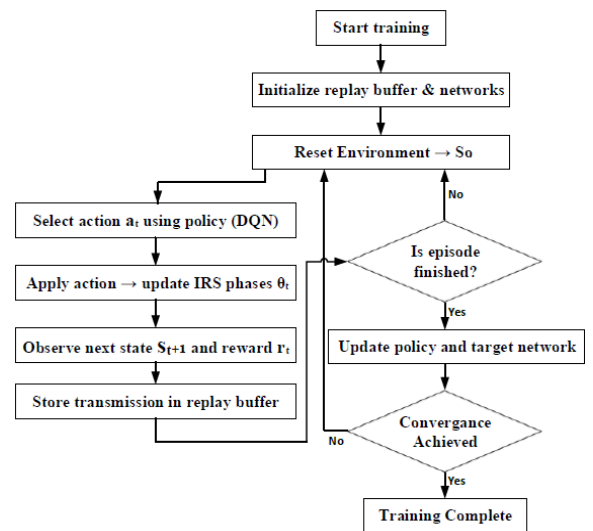


Figure 4. Flowchart of the proposed DRL training -based IRS phase-shift optimization algorithm

Figure 4, which displays the algorithm-level flowchart of the suggested DRL-based IRS phase-shift optimization technique, provides an illustration of the training framework.

Additionally, Figure 4 provides a comprehensive overview of the episode-based training loop, including the network update process, replay buffer operation, and agent-environment interaction.

3.2.4 Training workflow

The learning agent and the wireless environment interact iteratively during the training process of the suggested DQN-based IRS optimization framework. The starting state, which includes the applied IRS configuration and the current channel circumstances, is then obtained by resetting the environment. Using an ϵ -greedy strategy that strikes a balance between exploration and exploitation, the agent chooses an action based on this state. A discrete IRS phase-shift configuration is determined by the chosen action and applied to the reflecting surface. The wireless environment then changes into a new state and provides a scalar reward that represents the link performance that was attained. The replay buffer, which functions as the agent's experience memory, contains the resulting transition tuple (st, at, rt, s_{t+1}) . Gradient descent is used to update the online Q-network's parameters by randomly sampling mini-batches of previous transitions from the replay buffer. For stability of training, the target Q-network is periodically synced with the online network. This interaction loop is repeated for a number of time steps and training events until the learning process converges.

Pseudo-code presents the pseudocode of the suggested DQN-based IRS optimization algorithm, summarizing the whole training process and algorithmic flow. The pseudocode describes the following: initialization, agent environment interaction, experience storage, updates of the Q-network, and the convergence process.

Pseudo-code: Optimization Phase-Shift using DQN

Input: Target update interval C , mini-batch size M , replay buffer capacity β , discount factor γ , exploration rate ϵ , maximum steps T , maximum episodes E , IRS phase-shift codebook, Channel state information.

Output: Efficient IRS phase-shift configuration.

- 1: Initialize buffer β replay.
- 2: Initialize random weights w with online Q-network $Q(s, a; w)$.
- 3: Initialize weights $w \rightarrow w'$ with target Q-network $Q'(s, a; w')$.
- 4: Initialize θ_0 phase-shift vector of IRS.
- 5: For each episode $e = 1$ to E do
- 6: Achieve initial state S_0 and rearrange the environment.
- 7: For each step $= 1$ to T do
- 8: ϵ -greedy policy used to select action a_t .
- 9: Choose a random action: with probability ϵ .
- 10: Otherwise: achieve $a_t = \arg \max_a Q(s_t, a; w)$.
- 11: Apply the IRS phase-shift configuration of action a_t .
- 12: Obtain reward r_t and detect next state S_{t+1} .
- 13: In replay buffer β , stores the transition $(st, at, rt, st+1)$.
- 14: From buffer β , sample a mini-batch of M transitions.
- 15: Target determine: $y_t = r_t + \gamma \max_{a'} Q'(s_{t+1}, a'; w')$.
- 16: Loss determine: $L = (y_t - Q(s_t, a_t; w))^2$.
- 17: Minimizing loss L use to update parameters network

online w .

18: if $t \bmod C = 0$ then

19: Network target update $w \rightarrow w'$.

20: end if

21: end for

22: if met criterion convergence then

23: break.

24: end if

25: end for

26: Return improved IRS phase-shift formation.

In this formulation, the iterative learning mechanism by which the agent is able to converge to an effective IRS phase shift control policy is presented in a simple and repeatable manner.

3.2.5 Simulation setup

The simulation environment is optimized by using a DQN-based controller to determine the configuration of the IRS, which takes into account an IRS-assisted wireless communication link. The efficiency of the Numerical simulation is used to evaluate the suggested learning-based optimization method in a typical downlink IRS-assisted communication scenario. The IRS-DRL experiments' simulation parameters are shown in Table 5. The carrier frequency, bandwidth, transmit power, propagation lengths, IRS characteristics, and RL settings are all specified by these parameters. The selection of Communication related and learning related criteria is carefully selected in order to obtain a fair evaluation and simulate real deployment scenarios.

Table 5 shows an overview of the simulation parameters of the IRS-DRL architecture. Selected Settings define the wireless environment, IRS setup and channel circumstances used for the performance assessment. The setups are useful for testing the effectiveness of the proposed DRL-based phase-shift optimization technique. In addition, the simulation providing a solid basis for comparisons with the benchmark improvement methods that covered in the following section.

Table 5. Simulation parameters for the IRS-DRL experiment

Parameter	Value
Carrier frequency	2.4 GHz
Bandwidth	20 MHz
Noise variance	-90 dBm
BS transmit power	30 dBm
Distance (BS $\rightarrow$ IRS)	20 m
Distance (IRS $\rightarrow$ UE)	30 m
Number of IRS elements (N)	20, 40, 60, 80, 100, 120
IRS phase quantization levels (Q)	4
Rician K-factor	3 dB
Path loss exponent	2.2

4. RESULTS AND DISCUSSION

In this section, a detailed study performance of the suggested DQN-based IRS phase-shift optimization approach presents. The results are generated under actual wireless channel conditions and are benchmarked using a variety of baseline techniques, such as greedy selection, fixed-phase IRS, random IRS configuration, PSO, GA, and traditional beamforming. The primary topics of analysis are convergence behavior, spatial performance dispersion, SNR scaling with IRS size, spectral efficiency, achievable rate, and overall

comparative performance.

4.1 Convergence behavior of the DQN agent

The learning dynamics of the proposed DQN-based optimizer through the evolution of the training reward over 200 episodes is illustrated in Figure 5. During the initial training phase, the agent exhibits a significant increase in the reward value, which is suggestive of an exploration activity where many IRS phase-shift configurations have been sampled. The reward trajectory becomes smoother with training, and eventually evades to a stable plateau signifying transition from exploration to exploitation of the learnt strategy. The proposed DRL model can achieve near-best among the evaluated methods rewards very rapidly and steadily around 100-120 episodes.

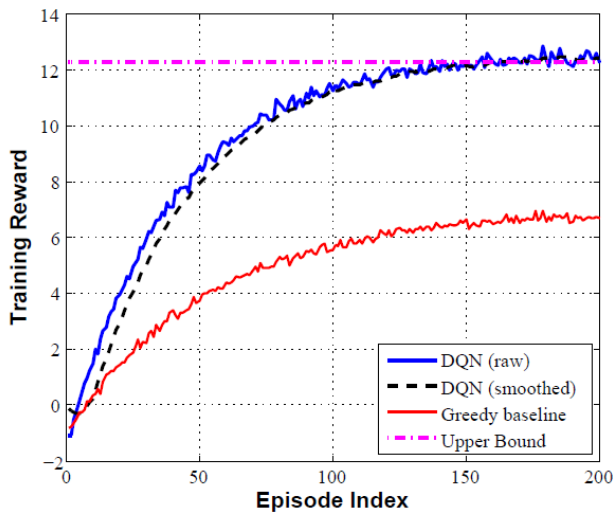


Figure 5. Convergence behavior of the suggested IRS optimization framework based on DQN

The DQN agent also finds a more efficient mapping between the IRS phase-shift configuration and the resultant received signal quality in comparison to deterministic baseline approaches and greedy selection.

This behavior demonstrates the learning process' resilience and stability. Furthermore, Figure 5 smoothed reward curve's consistent upward trend shows effective policy refinement with no discernible oscillations or divergence. The convergence behavior further validates the efficacy of the suggested learning-based optimization approach by indicating that the DQN agent can capture the intricate nonlinear link between the IRS phase-shift vector and the possible channel performance. This performance difference between the proposed DQN and the current DQN becomes more pronounced with the scaling of the IRS sizes due to the increased number of reflecting elements with which the phase can be optimized. Unlike a continuous-phase, the discrete phase codebook sets, thereby reducing complexity of the control signals available but at the expense of quantization constraint. The DQN is trained to identify good configurations in this set of discrete actions, leading to improved performance observed.

The proposed method needs an offline training phase requiring extra computational cost, but the trained network has a simple inference phase in the online application, which brings an appropriate balance for a practical application between training amount and the inference phase in the online

application. As shown in Figure 5, the training reward keeps going up over many episodes before it flattens out, which basically suggests the chosen ϵ -decay schedule allows enough discovery, and it does not push the system into premature convergence. However, for larger IRS arrays, a slower approach to ϵ -decay or more sophisticated DRL designs such as Double-DQN, Dueling-DQN and Actor-Critic may be required to maintain performance and to prevent its degradation as array size increases.

4.2 Achievable rate with various SNR situations

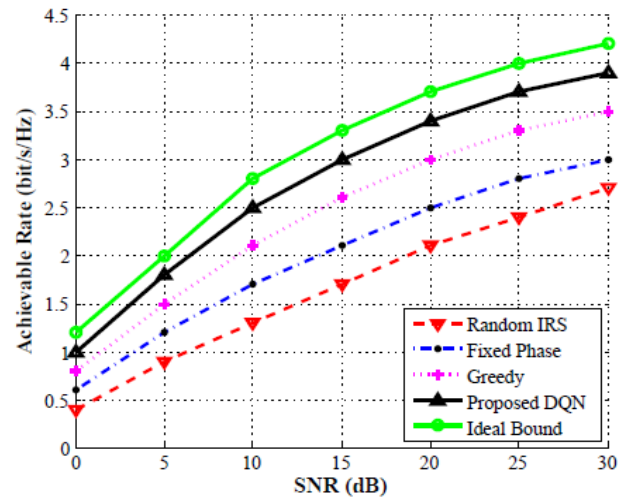
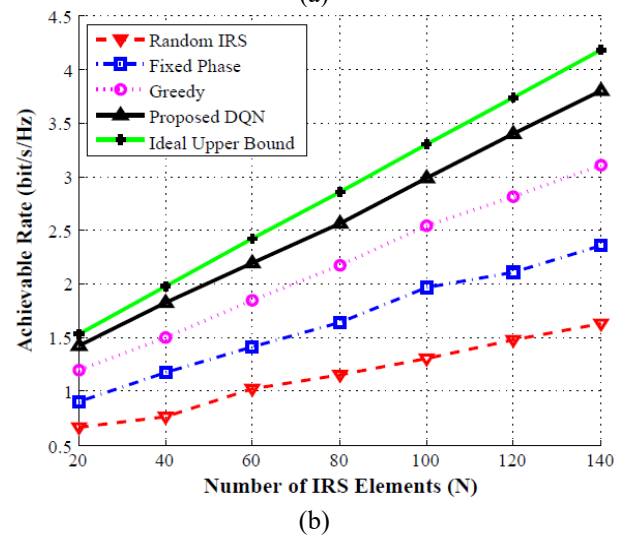


Figure 6. (a) SNR comparison of achievable rates for different optimization techniques. (b) Rate attainable in the quantity of pieces that reflect the IRS

The possible rate performance under various SNR levels is illustrated in Figure 6(a). The suggested DQN-assisted IRS design consistently outperforms all benchmark systems, including random IRS configuration, fixed-phase IRS, greedy selection, and conventional beamforming, throughout the whole operating SNR range from 0 to 30 dB. This performance benefit demonstrates how well the suggested learning-based phase-shift optimization technique works. In particular, over all SNR points, the DQN-based method continuously maintains a higher attainable rate. Because it lacks long-term optimization knowledge, the greedy method performs poorly even if it shows relatively quick convergence due to its local optimization nature. The trajectory of the suggested DQN

technique outperforms all other baselines and comes very near to the ideal upper bound curve, which acts as a theoretical performance ceiling. The suggested DRL technique produces more than 4.2 bits/s/Hz at an SNR of 30 dB, greatly closing the distance to the optimal upper bound. RL allows for adaptive IRS phase optimization capabilities that are not possible with simply heuristic-driven methods, as seen by the distinct divergence between the performance curves in Figure 6(a). The DQN agent successfully captures the intricate link between IRS phase designs and feasible data rates by utilizing long-term reward maximization. Figure 6(b) further explores the effect of IRS array size on the attainable rate by varying the number of reflecting elements over a broad range. The findings verify that a near-linear increase in the attainable rate due to increased passive beamforming gain occurs when the IRS size is increased. Interestingly, for every IRS size taken into consideration, the suggested DQN-based architecture performs better than other benchmark techniques, and the performance difference widens as the array size grows. This behavior illustrates the high-dimensional IRS scenario's improved performance under the considered simulation conditions. of the proposed learning-driven approach.

However, the performance gains of the greedy and fixed-phase processes are sublinear in terms of the number of pieces that reflect the IRS. When the IRS size relation to the number of IRS pieces increases, indicating limited use of the expanded beamforming region. For example, compared to the greedy and fixed-phase, the proposed DRL-based IRS optimization can achieve an achievable rate of approximately 35%-45% higher at $N = 120$. Combined, these results show that DRL is more readily adaptable and scalable to large scale wireless systems with IRS assistance.

4.3 SNR scaling in relation to IRS element

The received SNR as a function of the number of IRS reflecting elements is shown in Figure 7(a), offers more information about the scalability of the suggested system. The proposed optimization approach based on DRL always gives the highest SNR for all configurations analysed when the number of IRS elements increases. This behavior illustrates the adapting capacity of the DQN agent to different channel instantiations, and the utilization of the extra beamforming degrees of freedom offered by greater IRS array sizes. In particular, for $N = 120$, the suggested DQN-based method Figure 7(b), which shows the evolution of spectral efficiency throughout training episodes, provides further insight into the learning behavior in terms of spectral efficiency. There is a distinct two-phase pattern. The DQN agent quickly increases spectral efficiency during the first training phase as it investigates and finds efficient IRS setups. Reliable policy refinement is demonstrated by the curve's steady stabilization during training and convergence after about 150 sessions.

It is important to highlight that the training reward will converge around 100-120 episodes while the spectral efficiency will converge around 150 episodes. The convergence is the degree of change when the matching criterion metric changes by an insignificant amount after additional training trials. Heuristic baseline approaches, such as random selection, greedy search, PSO, and GA, on the other hand, show significantly lower spectral efficiency values with no progress over time since they are unable to improve judgments across episodes. Converging to a constant spectral efficiency of around 4.7 bits/s/Hz, the suggested DRL-based

strategy demonstrates its capacity to maximize link throughput beyond static or semi-static IRS designs by utilizing collected knowledge.

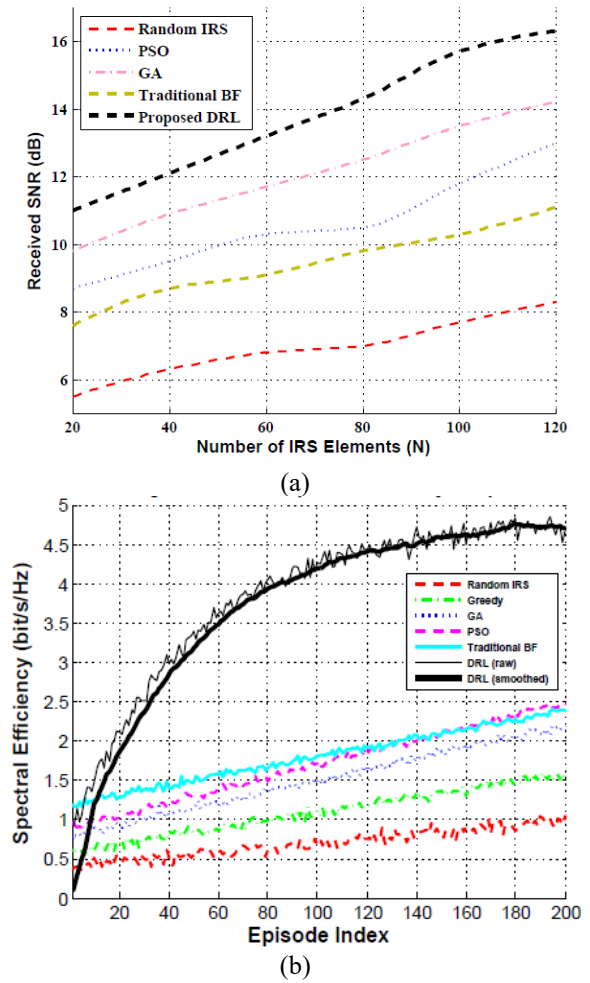


Figure 7. (a) Received SNR versus number of IRS reflecting elements for different optimization methods. (b) Spectral efficiency evolution over training episodes

4.4 Spatial distribution of signal power received

A two-dimensional heatmap of the received power throughout the area under consideration is shown in Figure 8, illustrating the spatial distribution of the received signal strength. Near the intended user position, a prominent high-power zone is readily visible, suggesting that the optimized IRS phase-shift configuration effectively focused the beam. By boosting constructive interference toward the intended user and reducing unwanted signal dispersion in nearby areas, the suggested DQN-based IRS optimization modifies the wireless propagation environment. This spatial behavior is in line with the quadratic array-gain principle, which states that concentrated energy focusing results from coherent phase alignment across IRS elements. The capacity of the DRL-optimized IRS to intelligently manage the wireless channel is confirmed by the development of a strong focus point around the user position. The potential of learning-based IRS solutions for real-world uses, such as improving indoor coverage, creating hotspots, and providing energy-focused wireless services, is highlighted by such spatial power control.

Denoted as 'Received SNR' in Figure 8, the spatial distribution of the received SNR distributed over the

considered UE region. The BS and IRS were placed at (-20 ,0) m and (0, 0) m while the selected UE position was located at (30, 0) m. The two-dimensional region indicated in the Figure is varied over the location of the UE. The simulation used a 2.4 GHz carrier frequency, 20MHz bandwidth, 30dBm BS transmit power and noise variance of -90dBm. These 120 reflecting elements were chosen from a set of four different quantization levels and controlled by the proposed optimization method DQN. Additional with x and y axes plot the UE's position in meters, and, the color scale plots the SNR received in dB.

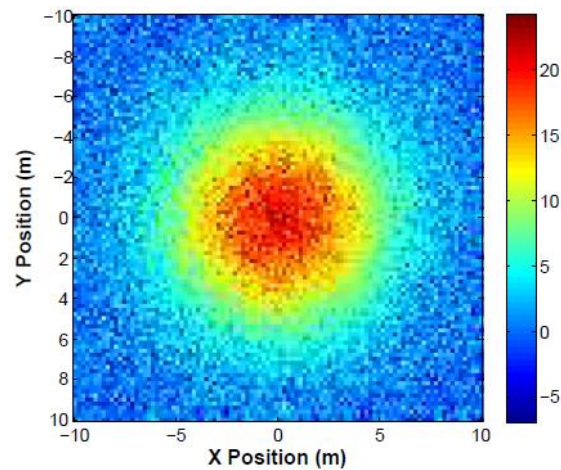


Figure 8. The received signal power spatial heatmap under the suggested DQN-based IRS modification

4.5 Combined metric comparison

A dual-axis bar chart that compares the received SNR and spectral efficiency attained by the IRS optimization strategies under consideration in order to jointly assess the major performance parameters is presented in Figure 9. The success of the suggested DQN-based technique in simultaneously improving signal quality and data throughput is demonstrated by the fact that it regularly beats all benchmark approaches across both parameters. In particular, the suggested DRL-based approach simultaneously reaches the greatest spectral efficiency of 4.73 bit/s/Hz and the highest received SNR of around 17.1 dB. Heuristic methods like fixed-phase IRS and random IRS configuration, on the other hand, show far lower values in both measures, indicating their poor capacity to take use of the IRS beamforming potential. The learning-driven optimization strategy's overall superiority over traditional methods is confirmed by this combined comparison. The random, fixed-phase, and traditional techniques produce around 8.2 dB, 14.2 dB, and 11.4 dB, respectively, whereas the suggested method obtains the greatest SNR of about 17.1 dB. The DRL-based strategy also achieves the highest value of 4.73 bit/s/Hz in spectral efficiency, whereas the competing methods achieve 1.17, 2.27, and 2.45 bit/s/Hz. The results confirm that the proposed DRL architecture is able to learn an optimal IRS phase-shift configuration, leading to increased spectrum efficiency and link quality.

The convergence behavior, spectral efficiency, and received SNR for the different IRS optimization methods are numerically compared in Table 6, which shows that the performance of all these schemes is good with minimal convergence error. The findings highlight numerous important findings. Firstly, the proposed DRL approach delivers the

greatest SNR by employing adaptive phase-shift selection to provide near-best improved coherent beamforming gain. Conversely, baseline methods such as traditional beamforming, PSO, GA, greedy search, and random selection cannot achieve comparable coherence, which results in decreased SNR performance. Secondly, in accordance with the logarithmic connection between spectral efficiency and SNR, the DRL-based approach achieves the highest spectral efficiency and SNR. It should be noted that the goal of the comparison with Figure 9 and Table 6 is different. The performance of the proposed DRL framework is compared with representative phase-selection strategies in Figure 9 and an extended comparison with conventional optimization-based algorithms is given in Table 6, such as the Greedy, GA and PSO methods. Thus, the spectral efficiency findings maintain the performance gap seen in the SNR domain, underscoring the efficacy of the suggested method. Lastly, the comparison of convergence speeds shows that the suggested DRL framework converges far more quickly than both traditional beamforming schemes and population-based optimization techniques like GA and PSO.

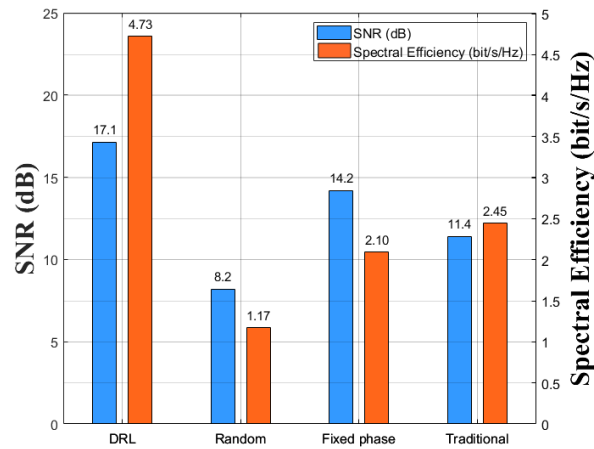


Figure 9. A bar chart that contrasts the spectral efficiency and received SNR for several IRS optimization techniques

Table 6. A comparison of the suggested DRL's performance with different algorithms

Method	SNR (dB)	Spectral Efficiency (bit/s/Hz)	Convergence Steps
Random IRS	8.2	1.17	300
Greedy	12.8	1.61	230
Traditional Beamforming	11.4	2.45	210
GA	13.8	2.85	170
PSO	14.9	3.25	150
Proposed DRL	17.1	4.73	120

All baseline methods were tested under the same channel, IRS configuration, and simulation parameters for a fair comparison with the proposed DQN. GA: population size of 30 and Number of iterations of 100, and PSO: swarm size of 30 and Number of iterations of 100. The Greedy method was done one-by-one for the available discrete choices of phase-shifts. The same 2-bit phase resolution ($Q = 4$) was used for all methods. This benefit results from the DQN agent's capacity to use the Bellman update process to learn an explicit mapping between channel states and ideal IRS phase

configurations.

The results reveal that the proposed IRS optimization using DQN translates to higher spectral efficiency, better SNR growth with the increase of IRS size, faster and more stable convergence as well as better achievable rates and better spatial signal focusing. The findings are in good agreement with those of DRL as an effective, scalable and reliable approach towards optimizing IRS in next-generation wireless communication systems. However, the estimate of the CSI can lead to some errors, which in actual deployments may affect both the maximum achievable spectral efficiency and the actual performance of the learned policy. Therefore, it is indeed a relevant path for further research, having taken a look at the robustness of the proposed approach under imperfect CSI.

The additional benefits of the proposed DRL-based IRS optimization approach are attributed to its ability of an adaptive IRS optimization for IR phase shifts in some degree, depending overall on the current wireless channel conditions. In this place learned phase layouts result in tighter or more focused spatial beams, and better constructive signal adding at the receiver. This turn results in an improved SNR at the receiver input, thereby enhancing the overall performance of the communication system in terms of the spectrum efficiencies and reliability.

5. CONCLUSION

A DRL-based framework for optimizing the phase-shifted design of IRS in wireless communication systems was presented in this study. Process decision based Markov chain used for problem control IRS framing and a DQN agent, the proposed methodology enables self-learning of virtually optimum IRS settings without the requirement for particular channel status models or manually developed optimization techniques. Numerous simulations demonstrate that the suggested DRL strategy performs noticeably better across a range of performance measures than existing optimization techniques using traditional beamforming, GA, PSO, greedy search, stable phase baseline, and randomized reflection.

In this study, results show that the DQN agent had the greatest received SNR (≈ 17.1 dB), the highest spectrum efficiency (≈ 4.73 bits/s/Hz), and the fastest convergence behavior. The learning trends and convergence curves confirmed the agent's stability and learning abilities, and the heatmap illustration showed how well the DRL-powered IRS directed energy toward the user position. Furthermore, under various benchmark conditions, IRS array sizes, and SNR levels, the results continuously improved.

These findings validate DRL's promise as a powerful tool for self-adaptive and computationally efficient IRS modification in subsequent-generation wireless networks. Future work is in progress to extend the framework to malfunctioning CSI and explore more complex DRL designs such as actor-critic, Dueling-DQN, and Double-DQN, techniques in order to enhance stability and sample efficiency.

REFERENCES

- [1] Wu, Q., Zhang, R. (2018). Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design. In 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, pp. 1-6. <https://doi.org/10.1109/GLOCOM.2018.8647620>
- [2] Wu, Q., Zhang, R. (2020). Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network. *IEEE Communications Magazine*, 58(1): 106-112. <https://doi.org/10.1109/MCOM.001.1900107>
- [3] Huang, C., Zappone, A., Alexandropoulos, G.C., Debbah, M., Yuen, C. (2019). Reconfigurable intelligent surfaces for energy efficiency in wireless communication. *IEEE Transactions on Wireless Communications*, 18(8): 4157-4170. <https://doi.org/10.1109/TWC.2019.2922609>
- [4] Gong, S., Lu, X., Hoang, D.T., et al. (2020). Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey. *IEEE Communications Surveys & Tutorials*, 22(4): 2283-2314. <https://doi.org/10.1109/COMST.2020.3004197>
- [5] Basar, E. (2020). Reconfigurable intelligent surface-based index modulation: A new beyond-MIMO paradigm for 6G. *IEEE Transactions on Communications*, 68(5): 3187-3196. <https://doi.org/10.1109/TCOMM.2020.2971486>
- [6] Han, Y., Tang, W., Jin, S., Wen, C., Ma, X. (2019). Large intelligent surface-assisted wireless communication exploiting statistical CSI. *IEEE Transactions on Vehicular Technology*, 68(8): 8238-8242. <https://doi.org/10.1109/TVT.2019.2923997>
- [7] Renzo, M.D., Debbah, M., Phan-Huy, D.T., et al. (2019). Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come. *EURASIP Journal on Wireless Communications and Networking*, 2019(1): 1-20. <https://doi.org/10.1186/s13638-019-1438-9>
- [8] Liu, Y., Liu, X., Mu, X., et al. (2021). Reconfigurable intelligent surfaces: Principles and opportunities. *IEEE Communications Surveys & Tutorials*, 23(3): 1546-1577. <https://doi.org/10.1109/COMST.2021.3077737>
- [9] Guo, H., Liang, Y.C., Chen, J., Larsson, E.G. (2019). Weighted sum-rate optimization for intelligent reflecting surface-enhanced wireless networks. In *Proceedings of the IEEE Global Communications Conference, Waikoloa, HI, USA*, pp. 1-6. <https://doi.org/10.1109/GLOBECOM38437.2019.9013288>
- [10] Zhang, S., Zhang, R. (2020). Capacity characterization for intelligent reflecting surface aided mimo communication. *IEEE Journal on Selected Areas in Communications*, 38(8): 1823-1838. <https://doi.org/10.1109/JSAC.2020.3000814>
- [11] Özdogan, Ö., Björnson, E. (2020). Deep learning-based phase reconfiguration for intelligent reflecting surfaces. *arXiv preprint arXiv:2009.13988*. <https://doi.org/10.48550/arXiv.2009.13988>
- [12] Zhou, G., Pan, C., Ren, H., Wang, K., Nallanathan, A. (2020). A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels. *IEEE Transactions on Signal Processing*, 68: 5092-5106. <https://doi.org/10.1109/TSP.2020.3019666>
- [13] He, Z.Q., Yuan, X. (2019). Cascaded channel estimation for large intelligent metasurface assisted massive MIMO. *IEEE wireless communications letters*, 9(2): 210-214. <https://doi.org/10.1109/LWC.2019.2948632>

- [14] Guo, H., Liang, Y.C., Chen, J., Larsson, E.G. (2020). Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks. *IEEE Transactions on Wireless Communications*, 19(5): 3064-3076. <https://doi.org/10.1109/TWC.2020.2970061>
- [15] Zhao, F., Chen, W., Liu, Z., Li, J., Wu, Q. (2023). Deep reinforcement learning-based intelligent reflecting surface optimization for TDD multi-user MIMO systems. *IEEE Wireless Communications Letters*, 12(11): 1951-1955. <https://doi.org/10.1109/LWC.2023.3301496>
- [16] Wang, Y., Ho, I.W.H., Zhang, S., Wang, Y. (2023). Intelligent reflecting surface enabled fingerprinting-based localization with deep reinforcement learning. *IEEE Transactions on Vehicular Technology*, 72(10): 13162-13172. <https://doi.org/10.1109/TVT.2023.3275581>
- [17] Taha, A., Zhang, Y., Mismar, F.B., Alkhateeb, A. (2020). Deep reinforcement learning for intelligent reflecting surfaces: Towards standalone operation. In *Proceedings of the IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Atlanta, GA, USA, pp. 1-5. <https://doi.org/10.1109/SPAWC48557.2020.9154301>
- [18] Yashvanth, L., Murthy, C.R. (2023). Performance analysis of intelligent reflecting surface assisted opportunistic communications. *IEEE Transactions on Signal Processing*, 71: 2056-2071. <https://doi.org/10.1109/TSP.2023.3265887>
- [19] Ye, J., Guo, S., Alouini, M.S. (2020). Joint reflecting and precoding designs for SER minimization in reconfigurable intelligent surfaces. *IEEE Transactions on Wireless Communications*, 19(8): 5561-5574. <https://doi.org/10.1109/TWC.2020.2994455>
- [20] Hao, X., Feng, Z., Yang, S., Wang, M., Jiao, L. (2023). Automatic modulation classification via meta-learning. *IEEE Internet of Things Journal*, 14: 12276-12292.
- [21] Abdallah, A., Celik, A., Mansour, M.M., Eltawil, A.M. (2025). Multi-agent DRL for distributed codebook design in RIS-aided cell-free massive MIMO Networks. *IEEE Transactions on Communications*, 73(5): 3283-3297. <https://doi.org/10.1109/TCOMM.2024.3483041>
- [22] You, C., Zheng, B., Zhang, R. (2020). Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement. *IEEE Journal on Selected Areas in Communications*, 38(11): 2604-2620. <https://doi.org/10.1109/JSAC.2020.3007056>
- [23] Basar, E., Di Renzo, M., De Rosny, J., Debbah, M., Alouini, M.S., Zhang, R. (2019). Wireless communications through reconfigurable intelligent surfaces. *IEEE access*, 7: 116753-116773. <https://doi.org/10.1109/ACCESS.2019.2935192>
- [24] Hu, X., Zhong, C., Alouini, M.-S., Zhang, Z. (2021). Robust design for IRS-aided communication systems with user location uncertainty. *IEEE Wireless Communications Letters*, 10(1): 63-67. <https://doi.org/10.1109/LWC.2020.3020850>
- [25] Jiang, W., Ai, B., Li, M., Wu, W., Shen, X. (2023). Average age-of-information minimization in aerial IRS-assisted data delivery. *IEEE Internet of Things Journal*, 10(17): 15133-15146.
- [26] Illi, E., Bazzi, A., Qaraqe, M., Ghayeb, A. (2025). On the secrecy-sensing optimization of RIS-assisted full-duplex integrated sensing and communication network. *IEEE Transactions on Wireless Communications*, 25: 9530-9547. <https://doi.org/10.1109/TWC.2025.3645048>
- [27] Elbir, A.M., Mishra, K.V. (2025). A survey of deep learning architectures for intelligent reflecting surfaces. *IEEE Communications Standards Magazine*, 10(2): 12-19. <https://doi.org/10.1109/MCOMSTD.2025.3616430>
- [28] Volakaki, M.G., Papaioannou, G., Vouyioukas, D. (2026). Joint wave beamforming and RIS phase shift optimization in multi-user group networks under UAV jamming. *IEEE Open Journal of the Communications Society*, 7: 2034-2055. <https://doi.org/10.1109/OJCOMS.2026.3668366>
- [29] Feng, K., Wang, Q., Li, X., Wen, C.K. (2020). Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems. *IEEE Wireless Communications Letters*, 9(5): 745-749. <https://doi.org/10.1109/LWC.2020.2969167>
- [30] Wang, L., Wang, K., Pan, C., Aslam, N. (2022). Joint trajectory and passive beamforming design for intelligent reflecting surface-aided UAV communications: A deep reinforcement learning approach. *IEEE Transactions on Mobile Computing*, 22(11): 6543-6553. <https://doi.org/10.1109/TMC.2022.3200998>
- [31] Wang, W., Zhang, W. (2022). Intelligent reflecting surface configurations for smart radio using deep reinforcement learning. *IEEE Journal on Selected Areas in Communications*, 40(8): 2335-2346. <https://doi.org/10.1109/JSAC.2022.3180787>
- [32] Shehab, M., Ciftler, B.S., Khatlab, T., Abdallah, M.M., Trinchero, D. (2022). Deep reinforcement learning powered IRS-assisted downlink NOMA. *IEEE Open Journal of the Communications Society*, 3: 729-739. <https://doi.org/10.1109/OJCOMS.2022.3165590>
- [33] Evmorfos, S., Petropulu, A.P., Poor, H.V. (2022). Deep reinforcement learning for IRS phase shift design in spatiotemporally correlated environments. *arXiv preprint arXiv:2211.09726*. <https://doi.org/10.48550/arXiv.2211.09726>
- [34] Vishnoi, V., Consul, P., Budhiraja, I., Gupta, S., Kumar, N. (2023). Deep reinforcement learning based energy consumption minimization for intelligent reflecting surfaces assisted D2D users underlaying UAV network. In *Proceedings of the IEEE Conference on Computer Communications Workshops*, Hoboken, NJ, USA, pp. 1-6. <https://doi.org/10.1109/INFOCOMWKSHPS57453.2023.10226057>
- [35] Ikeagu, K., Ding, Y., Song, C., Khandaker, M.R. (2023). Intelligent reflecting surface optimization for MIMO communication using deep reinforcement learning. In *2023 31st Telecommunications Forum (TELFOR)*, Belgrade, Serbia, pp. 1-4. <https://doi.org/10.1109/TELFOR59449.2023.10372753>