Adaptive Multi-Mode DRL Framework for Intelligent RIS-Assisted Anti-Jamming in 6G Wireless Networks

Le Hoang Hiep, and Huu-Huy Ngo*

***Abstract*—This paper proposes an adaptive multi-mode Deep Reinforcement Learning (DRL) framework for intelligent RIS-assisted anti-jamming communication in dynamic 6G wireless networks. The proposed Framework jointly integrates RIS beamforming, channel hopping, and transmit power adaptation through a DRL-Driven decision engine capable of dynamically responding to varying interference conditions and channel fluctuations. To improve deployment realism, practical constraints including imperfect Channel State Information (CSI), finite-resolution RIS phase quantization, reflection loss, control delay, and user mobility are incorporated into the system model. The anti-jamming problem is formulated as a Markov decision process and solved using DQN, PPO, and SAC algorithms. Extensive simulations are conducted using MATLAB-based wireless channel modeling and Python-based DRL training platforms. Simulation results demonstrate that the proposed framework achieves approximately 25%–40% higher throughput and 18%–35% SINR improvement compared with conventional anti-jamming approaches. Moreover, the proposed scheme maintains stable communication performance under strong jamming power, CSI uncertainty, and high-mobility scenarios. Statistical evaluations over 20 independent random seeds further confirm the robustness and reproducibility of the proposed framework.**

***Index Terms*—6G wireless networks, Anti-jamming communications, Reconfigurable intelligent surfaces (RIS), Deep Reinforcement Learning (DRL), Adaptive multi-mode optimization.**

I. Introduction

Sixth-generation (6G) wireless networks are expected to provide ultra-reliable and intelligent communication services for emerging applications such as massive Internet of Things (IoT), autonomous systems, digital twins, and immersive extended reality [1], [2]. Compared with previous wireless generations, 6G systems are envisioned to support significantly higher spectral efficiency, ultra-low latency, and dense device connectivity under highly dynamic communication environments [3], [4]. However, the open and shared nature of wireless channels also makes future 6G networks increasingly vulnerable to intentional interference and jamming attacks, which can severely degrade communication reliability, throughput performance, and quality of service [5].

Recently, reconfigurable intelligent surfaces (RISs) have emerged as a promising technology for enhancing wireless propagation environments through programmable electromagnetic reflection control [6], [7]. By intelligently adjusting RIS reflection coefficients, wireless signals can be constructively combined at desired receivers while interference can be suppressed [8], [9]. Owing to these advantages, RIS-assisted communication has attracted considerable attention for secure transmission, interference mitigation, and spectrum efficiency enhancement in next-generation wireless systems [10], [11].

At the same time, deep reinforcement learning (DRL) has demonstrated strong capability in solving dynamic wireless optimization problems without requiring accurate mathematical models of complex environments [12], [13]. Recent studies have applied DRL to anti-jamming communications through adaptive channel selection, transmit power control, and resource allocation strategies [14]–[16]. In the broader context of wireless sensor networks, link selection mechanisms have also been investigated to enhance communication reliability under dynamic channel conditions [17]. In parallel, several RIS-assisted anti-jamming schemes have been investigated to improve signal robustness under hostile interference conditions [18]–[20].

Nevertheless, a careful examination of recent literature reveals several important limitations. First, most existing DRL-based anti-jamming approaches optimize only a single communication dimension—such as channel hopping or power adaptation—which limits their adaptability under heterogeneous jamming environments [21], [22]. Second, current RIS-assisted schemes typically assume idealized deployment conditions, including perfect channel state information (CSI), continuous RIS phase shifts, negligible hardware impairments, and static wireless environments [23], [24]. Third, although several recent studies have separately investigated RIS optimization, DRL-based anti-jamming, or adaptive resource allocation, very limited work has jointly integrated RIS beamforming, channel selection, and transmit power adaptation into a unified multi-mode decision framework [25], [26]. In particular, most reported schemes rely on fixed anti-jamming policies and do not support dynamic mode switching according to jammer behavior, interference intensity, or channel variations [27], [28].

Manuscript received April 7, 2026; revised May 13, 2026. Date of publication July 8, 2026. Date of current version July 8, 2026. The associate editor prof. Adriana Lipovac has been coordinating the review of this manuscript and approved it for publication.

Authors are with the Thai Nguyen University of Information and Communication Technology, Thai Nguyen, Vietnam (e-mails: lhhiep@ictu.edu.vn, nhhuy@ictu.edu.vn).

Digital Object Identifier (DOI): 10.24138/jcomss-2026-0077

*Corresponding author

Furthermore, recent survey studies on RIS-assisted anti-jamming communications have highlighted that practical deployment constraints—including finite-resolution RIS phase quantization, reflection loss, control delay, imperfect CSI, and mobility-aware channel dynamics—remain insufficiently investigated in existing DRL-based frameworks [29]. Several multi-agent and cooperative anti-jamming schemes have been proposed for UAV swarm scenarios [30], yet the integration of practical RIS hardware impairments with adaptive DRL decision making in terrestrial 6G settings remains largely unexplored. Consequently, the robustness and reproducibility of many reported performance gains remain difficult to verify under realistic 6G deployment scenarios.

Motivated by these observations, this paper proposes an adaptive multi-mode DRL framework for RIS-assisted anti-jamming communication in dynamic 6G wireless networks. Unlike conventional single-strategy approaches, the proposed framework dynamically switches among RIS beamforming, channel hopping, and transmit power adaptation according to real-time wireless conditions and jammer activities. Practical deployment factors, including imperfect CSI, RIS phase quantization, reflection loss, control delay, and mobility effects, are explicitly incorporated into the system model to improve deployment realism.

The main contributions of this paper are summarized as follows:

- A practical RIS-assisted anti-jamming system model is developed by incorporating imperfect CSI, finite-resolution RIS phase quantization, reflection loss, control delay, and mobility-aware wireless channels.
- An adaptive multi-mode DRL framework is proposed to jointly optimize RIS beamforming, channel selection, and transmit power adaptation under dynamic jamming environments.
- A unified action-space design is introduced to support both discrete and continuous anti-jamming decisions using DQN, PPO, and SAC algorithms.
- Extensive simulations and statistical evaluations over 20 independent random seeds demonstrate that the proposed framework achieves approximately 25%–40% higher throughput and 18%–35% SINR improvement compared with benchmark schemes under practical 6G deployment conditions.

Table I summarizes the main differences between recent RIS-assisted anti-jamming approaches and the proposed framework. Unlike existing methods, the proposed framework jointly integrates adaptive multi-mode DRL decision making, practical RIS deployment constraints, and statistical robustness evaluation under dynamic wireless environments.

As observed from Table I, existing approaches generally focus on isolated optimization objectives or simplified wireless assumptions. In contrast, the proposed framework simultaneously considers adaptive anti-jamming mode switching, practical RIS impairments, mobility-aware environments, and statistical reproducibility, which remain insufficiently addressed in current literature.

The remainder of this paper is organized as follows. Section II presents the system model and problem formulation. Section III introduces the proposed adaptive multi-mode DRL framework and training methodology. Section IV provides simulation results and performance evaluations under practical wireless constraints. Finally, Section V concludes the paper and discusses future research directions.

II. System Model

A. Network Architecture

Consider a RIS-assisted wireless communication system operating in the presence of intentional jamming attacks, where a base station (BS) communicates with a legitimate user (LU) under the assistance of a reconfigurable intelligent surface (RIS). The considered network consists of four major components, namely the BS, the LU, the jammer (J), and the RIS controller.

The RIS is composed of N nearly passive reflecting elements capable of dynamically adjusting the phase response of incident electromagnetic waves. Unlike ideal RIS assumptions commonly adopted in existing studies, practical hardware limitations are considered in this work. Specifically, each RIS element employs finite-resolution phase shifters with b-bit quantization. Therefore, the phase shift of the n-th RIS element is constrained as

$$\theta_n \in \left\{0, \frac{2\pi}{2^b}, \ldots, \frac{2\pi(2^b-1)}{2^b}\right\}. \tag{1}$$

In addition, practical reflection loss caused by RIS hardware impairments is incorporated through a reflection amplitude coefficient $\rho \in (0, 1]$. Consequently, the RIS reflection matrix is modeled as

$$\mathbf{\Phi} = \mathrm{diag}\left(\rho e^{j\theta_1}, \rho e^{j\theta_2}, \ldots, \rho e^{j\theta_N}\right). \tag{2}$$

The BS communicates with the LU through both direct and RIS-reflected propagation links. By properly adjusting the RIS phase shifts, the reflected signals can be constructively combined at the receiver to enhance the desired signal power while suppressing jamming interference.

Furthermore, a practical RIS control delay is considered. Let τ_d denote the RIS reconfiguration delay caused by controller processing and signaling overhead. Accordingly, the applied RIS configuration at time slot t becomes $\mathbf{\Phi}(t - \tau_d)$, which better reflects practical RIS-assisted wireless deployments in future 6G systems.

B. Channel Model

The received signal at the legitimate user consists of both direct and RIS-assisted transmission components under dynamic wireless fading conditions.

Let h_d denote the complex channel coefficient of the direct BS-to-user link. The RIS-assisted transmission path consists of the BS-to-RIS channel $\mathbf{h}_{BR} \in \mathbb{C}^{N\times 1}$ and the RIS-to-user channel $\mathbf{h}_{RU} \in \mathbb{C}^{N\times 1}$.

Considering practical wireless environments, imperfect channel state information (CSI) is incorporated into the proposed framework. The estimated channels are modeled as

TABLE I
COMPARISON WITH EXISTING RIS-ASSISTED ANTI-JAMMING APPROACHES.

Reference	*RIS*	*DRL*	*Multi-Mode*	*Joint Optimization*	*Imperfect CSI*	*RIS Quantization*	*Mobility-Aware*	*Statistical Validation*
[14], [31]	✓	×	×	Partial	Partial	×	×	×
[18], [32]	✓	×	×	Partial	×	×	×	×
[25], [21]	×	✓	×	Partial	×	×	×	×
[19], [28]	✓	✓	×	Partial	×	×	×	×
[24], [33]	✓	×	×	✓	Partial	×	×	×
Proposed Framework	✓	✓	✓	✓	✓	✓	✓	✓

$$\hat{\mathbf{h}}_{BR} = \mathbf{h}_{BR} + \Delta\mathbf{h}_{BR}, \tag{3}$$

$$\hat{\mathbf{h}}_{RU} = \mathbf{h}_{RU} + \Delta\mathbf{h}_{RU}, \tag{4}$$

where $\Delta\mathbf{h}_{BR}$ and $\Delta\mathbf{h}_{RU}$ represent channel estimation errors following complex Gaussian distributions.

The equivalent RIS-assisted cascaded channel is expressed as

$$h_{RIS} = \hat{\mathbf{h}}_{RU}^{H}\boldsymbol{\Phi}\hat{\mathbf{h}}_{BR}. \tag{5}$$

To capture user mobility and time-varying propagation effects, Doppler-induced fading is additionally considered. The Doppler frequency is given by

$$f_D = \frac{v f_c}{c}, \tag{6}$$

where v denotes the user velocity, f_c is the carrier frequency, and c is the speed of light.

Assuming the BS transmits signal x with power P_t, where $\mathbb{E}[|x|^2] = 1$, the received signal at the LU can be expressed as

$$y = \left(h_d + \hat{\mathbf{h}}_{RU}^{H}\boldsymbol{\Phi}\hat{\mathbf{h}}_{BR}\right)\sqrt{P_t}x + i_J + n, \tag{7}$$

where i_J denotes the jamming interference and $n \sim \mathcal{CN}(0, \sigma^2)$ represents additive white Gaussian noise.

C. Jamming Model

In this work, two representative jamming strategies are considered.

First, the *reactive jammer* transmits interference signals only when legitimate transmission activity is detected. The received jamming signal at the user can be modeled as

$$i_J = h_J\sqrt{P_J}s \tag{8}$$

where h_J is the channel coefficient from the jammer to the user, P_J is the jamming power, and s is the jamming signal satisfying $\mathbb{E}[|s|^2] = 1$.

Second, the *random jammer* transmits interference signals randomly over time without observing channel activity. Although less intelligent than reactive jamming, it can still increase the background interference level and degrade communication reliability. The considered system model is shown in Fig. 1.

D. Problem Formulation

The objective of the considered RIS-assisted anti-jamming system is to jointly maximize communication reliability and energy efficiency while mitigating the impact of malicious interference under practical wireless constraints.

Based on the received signal model, the received signal-to-interference-plus-noise ratio (SINR) at the legitimate user is expressed as

$$\text{SINR} = \frac{P_t\left|h_d + \hat{\mathbf{h}}_{RU}^{H}\boldsymbol{\Phi}\hat{\mathbf{h}}_{BR}\right|^2}{P_J|h_J|^2 + \sigma^2}. \tag{9}$$

According to Shannon theory, the achievable throughput is defined as

$$R = B\log_2(1 + \text{SINR}), \tag{10}$$

where B denotes the system bandwidth.

In addition to throughput maximization, practical 6G communication systems must also consider latency and energy consumption constraints. The total system energy consumption is modeled as

$$E_{total} = P_t + N \times P_{RIS} + P_c, \tag{11}$$

where N denotes the number of reflecting elements of the RIS. P_{RIS} represents the power consumed by each RIS element and P_c denotes the circuit power consumption.

The overall communication delay is modeled as

$$D_{total} = D_{tx} + D_{RIS} + D_{proc}, \tag{12}$$

where D_{tx}, D_{RIS}, and D_{proc} denote transmission delay, RIS control delay, and DRL processing delay, respectively.

Accordingly, the anti-jamming optimization problem is formulated as

$$\mathbf{P1}: \max_{\boldsymbol{\Phi}, P_t, c_t} \quad \lambda_1 R - \lambda_2 E_{total} - \lambda_3 D_{total} \tag{13}$$

$$\text{s.t.} \quad 0 \leq P_t \leq P_{\max}, \tag{14}$$

$$\theta_n \in \left\{0, \frac{2\pi}{2^b}, \ldots, \frac{2\pi(2^b - 1)}{2^b}\right\}, \tag{15}$$

$$D_{total} \leq D_{\max}, \tag{16}$$

$$E_{total} \leq E_{\max}, \tag{17}$$

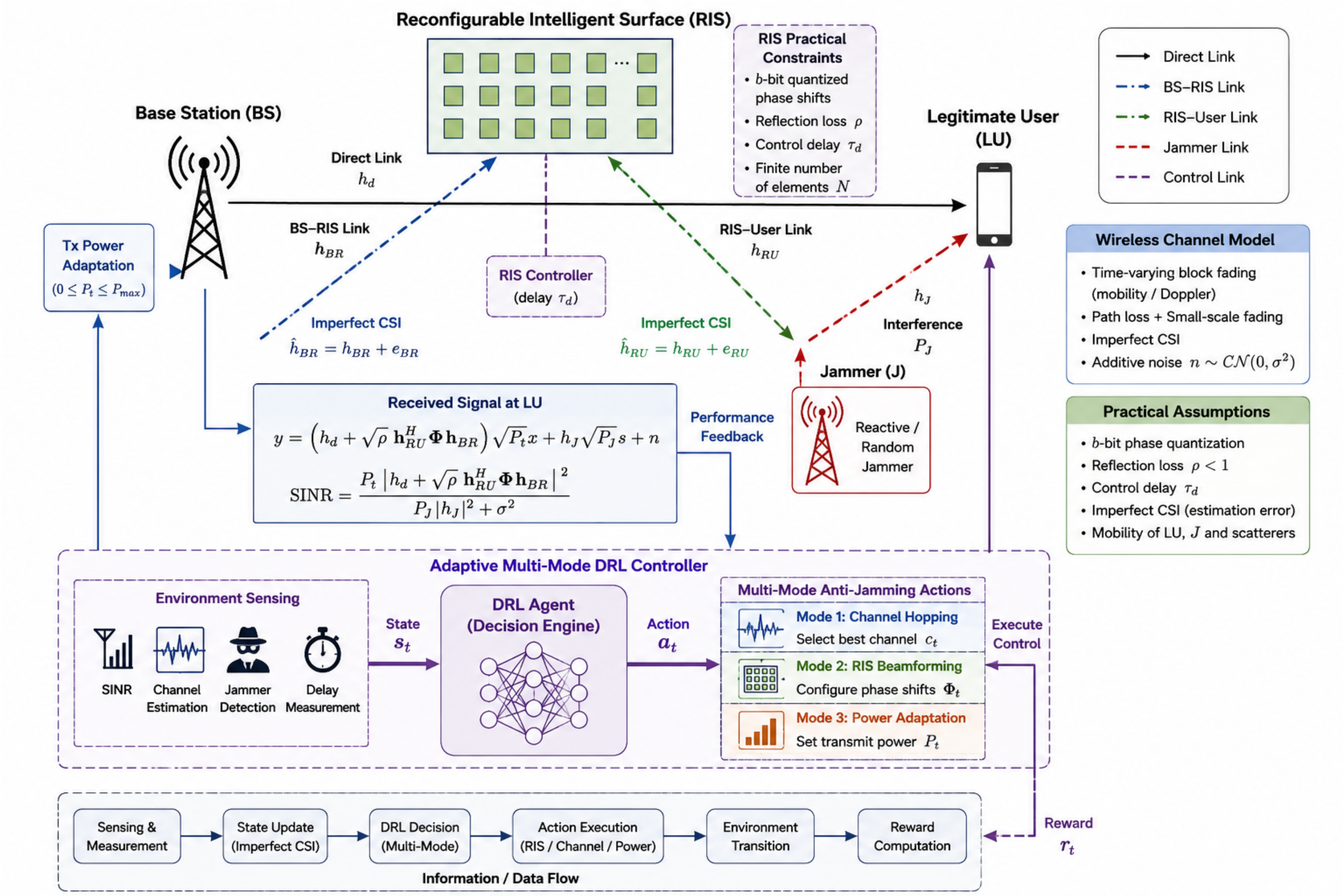

Fig. 1. Overall architecture of the practical RIS-assisted adaptive multi-mode DRL anti-jamming communication system under dynamic wireless environments.

where λ_1, λ_2, and λ_3 are weighting coefficients balancing throughput, energy efficiency, and latency performance, respectively.

The optimization problem is highly non-convex due to the coupled RIS reflection coefficients, discrete channel selection variables, and dynamic wireless environment uncertainties. Therefore, a DRL-based adaptive multi-mode optimization framework is proposed to learn near-optimal anti-jamming strategies in real-time wireless environments.

III. Proposed Adaptive Multi-Mode DRL Framework

A. Framework Overview

To address dynamic jamming attacks in 6G wireless networks, this paper proposes an adaptive multi-mode deep reinforcement learning (DRL) framework that jointly optimizes RIS configuration and communication parameters. The proposed framework consists of three main modules: environment sensing, DRL decision engine, and multi-mode anti-jamming control.

The *environment sensing* module collects wireless environment information, including channel conditions, interference levels, and jammer activity. Based on these observations, the system state at time slot t is defined as

$$s_t = \{\mathrm{SINR}_t, H_t, J_t\} \tag{18}$$

where SINR_t denotes the signal-to-interference-plus-noise ratio, H_t represents the channel condition, and J_t indicates the activity level of the jammer.

The *DRL decision engine* acts as an intelligent agent that determines optimal control actions based on the observed state. At each time step, the agent selects an action a_t to maximize the long-term cumulative reward.

The *multi-mode anti-jamming control* module executes the selected strategy, including channel hopping, RIS beamforming, or transmit power adaptation.

The interaction between the agent and the wireless environment can be modeled as a Markov Decision Process (MDP), where the objective is to learn an optimal policy π^*:

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right] \tag{19}$$

where γ is the discount factor and r_t denotes the reward at time t.

The overall architecture of the proposed adaptive multi-mode DRL framework is illustrated in Fig. 2. The framework consists of environment sensing, a DRL-based decision engine, and a multi-mode anti-jamming controller. The DRL

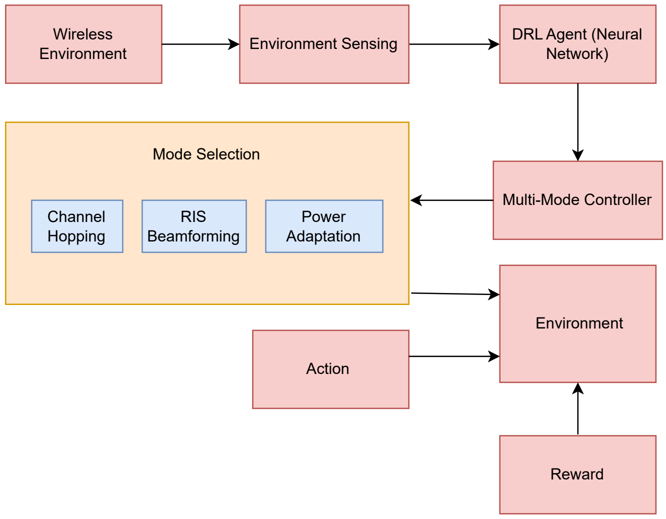

Fig. 2. Overall architecture of the proposed adaptive multi-mode deep reinforcement learning (DRL) framework for RIS-assisted anti-jamming wireless networks.

agent observes the wireless environment state and dynamically selects appropriate anti-jamming strategies, including channel hopping, RIS beamforming, and power adaptation.

B. Multi-Mode Anti-Jamming Strategy

Unlike conventional anti-jamming approaches that rely on a single mitigation technique, the proposed framework introduces a multi-mode strategy that dynamically selects different countermeasures according to the wireless environment.

1) Mode 1: Channel Hopping: In this mode, the transmitter switches to a cleaner frequency channel to avoid jamming interference. Let the available channel set be

$$\mathcal{C} = \{c_1, c_2, ..., c_K\}. \tag{20}$$

The optimal channel is selected as

$$c_t^* = \arg\max_{c_k \in \mathcal{C}} \mathrm{SINR}(c_k). \tag{21}$$

This strategy is effective when the jammer concentrates interference on specific channels.

2) Mode 2: RIS Beamforming: In this mode, the phase shifts of RIS elements are adjusted to strengthen the desired signal and mitigate interference. The RIS reflection matrix can be expressed as

$$\mathbf{\Phi} = \mathrm{diag}(e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_N}), \tag{22}$$

where θ_n denotes the phase shift of the n-th RIS element.

The RIS configuration aims to maximize the effective channel gain

$$\max_{\theta_n} \left| h_d + \mathbf{h}_{RU}^H \mathbf{\Phi} \mathbf{h}_{BR} \right|. \tag{23}$$

By properly adjusting the phase shifts, the RIS can create constructive signal combining at the legitimate receiver.

3) Mode 3: Power Adaptation: In this mode, the BS dynamically adjusts the transmit power to balance communication performance and energy efficiency. The transmit power is constrained by

$$0 \le P_t \le P_{\max}. \tag{24}$$

The resulting SINR can be expressed as

$$\mathrm{SINR} = \frac{P_t |h_{\mathrm{eff}}|^2}{P_J |h_J|^2 + \sigma^2}, \tag{25}$$

where h_{eff} denotes the effective channel including the RIS-reflected path.

C. DRL Model Design

The anti-jamming decision process is formulated using a DRL agent interacting with the wireless environment.

1) State Space: The system state at time t is defined as

$$s_t = [\mathrm{SINR}_t, H_t, J_t]. \tag{26}$$

Here, SINR_t indicates link quality, H_t represents the channel state, and J_t characterizes jammer activity.

2) Action Space: The proposed framework involves a hybrid action space consisting of both discrete and continuous control variables.

At each time slot, the DRL agent selects an action vector defined as

$$a_t = [m_t, c_t, \theta_1, \theta_2, \dots, \theta_N, P_t], \tag{27}$$

where:

- m_t denotes the anti-jamming mode selection,
- c_t represents the selected communication channel,
- θ_n denotes the RIS phase shift of the n-th reflecting element,
- P_t represents the BS transmit power.

The action variables include both discrete and continuous components. Specifically, mode selection and channel hopping are discrete actions, whereas RIS beamforming and transmit power adaptation are continuous control variables.

For the DQN-based implementation, continuous variables are discretized into finite action sets to enable Q-learning optimization. In particular, RIS phase shifts are quantized using finite-resolution phase control, while the transmission power is selected from a predefined discrete power set.

In contrast, PPO and SAC directly operate over continuous action spaces, allowing smoother RIS beamforming and power adaptation decisions under dynamic wireless environments.

3) Reward Function: The reward function is designed to encourage high throughput while suppressing interference. It is defined as

$$r_t = \alpha R_t - \beta I_t, \tag{28}$$

where R_t denotes system throughput, I_t represents interference power, and α and β are weighting coefficients.

4) Computational Complexity and Implementation Overhead: The computational complexity of the proposed DRL framework depends on the adopted learning paradigm and the hybrid action space. For DQN-based learning, the complexity scales as $\mathcal{O}(|\mathcal{A}| \cdot N_{\text{NN}})$, where $|\mathcal{A}|$ denotes the action space size and N_{NN} is the number of neural network parameters. Policy-gradient methods such as PPO and SAC introduce additional overhead due to gradient estimation and iterative policy updates. The action space grows with the number of RIS elements (32–128 in Table II), leading to increased training complexity due to the enlarged state–action space. Furthermore, the use of a replay buffer of size 10^5 adds moderate memory overhead. Nevertheless, the inference phase remains lightweight, requiring only a forward pass through a compact neural network, resulting in millisecond-level latency that is compatible with RIS control delays. Therefore, while training is computationally intensive, it can be executed offline, and the proposed framework remains feasible for real-time 6G deployments. Unlike Section III-F, which focuses on the training complexity, this subsection discusses both computational overhead and practical implementation aspects, including inference latency and memory requirements.

D. Training Algorithm

To learn the optimal anti-jamming policy, three representative DRL algorithms are implemented and compared, including Deep Q-Network (DQN), Proximal Policy Optimization (PPO), and Soft Actor-Critic (SAC).

The DRL agent employs a fully connected neural network consisting of three hidden layers with 128 neurons per layer and ReLU activation functions. The Adam optimizer is adopted with a learning rate of 10^{-4}. For DQN training, an experience replay buffer with capacity 10^5 is utilized to stabilize learning performance, while the mini-batch size is set to 64.

For DQN, the action-value function $Q(s,a)$ is approximated through a deep neural network. The corresponding loss function is expressed as

$$L(\theta) = \mathbb{E}\left[\left(r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta)\right)^2\right]. \quad (29)$$

For PPO, the policy network is updated by maximizing the clipped surrogate objective

$$L^{PPO} = \mathbb{E}\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]. \quad (30)$$

Unlike DQN and PPO, SAC adopts a maximum entropy reinforcement learning framework to improve exploration capability in highly dynamic jamming environments. Its objective function is defined as

$$J(\pi) = \mathbb{E}\left[\sum_t r_t + \alpha \mathcal{H}(\pi(\cdot|s_t))\right]. \quad (31)$$

Training convergence is assumed when the moving average reward variation becomes sufficiently small over consecutive training episodes.

Algorithm 1: Adaptive Multi-Mode DRL Anti-Jamming Framework

Initialize DRL network parameters θ;
Initialize target network parameters θ^{-rock};
Initialize replay buffer $\mathcal{D}$ with capacity 10^5;
Set learning rate η, discount factor γ, and mini-batch size B;
for *each training episode* **do**
  Reset wireless environment and RIS controller;
  Observe initial state s_t;
  while *episode not terminated* **do**
    Estimate wireless environment state:
    $s_t = \{\text{SINR}_t, \hat{H}_t, J_t, E_t, D_t\}$;
    Observe imperfect CSI and jammer activity;
    Select action $a_t = \{m_t, c_t, \mathbf{\Phi}_t, P_t\}$ using policy $\pi(s_t)$;
    `// Determine anti-jamming mode:`
    • Mode 1: Channel hopping
    • Mode 2: RIS beamforming
    • Mode 3: Power adaptation ;
    Quantize RIS phase shifts: $\theta_n \in \left\{0, \frac{2\pi}{2^b}, \ldots, \frac{2\pi(2^b-1)}{2^b}\right\}$;
    `// Execute selected action:`
    • Update communication channel c_t
    • Configure RIS reflection matrix $\mathbf{\Phi}_t$
    • Adjust BS transmit power P_t ;
    Apply RIS control delay τ_d;
    Observe next state s_{t+1};
    Compute reward: $r_t = \lambda_1 R_t - \lambda_2 E_t - \lambda_3 D_t - \lambda_4 I_t$;
    Store transition (s_t, a_t, r_t, s_{t+1}) into replay buffer $\mathcal{D}$;
    Sample mini-batch from $\mathcal{D}$;
    `// Update DRL parameters using:`
    • DQN for discretized action optimization,
    • PPO for stable policy-gradient learning,
    • SAC for continuous RIS optimization ;
    Update target network parameters periodically;
    Set $s_t \leftarrow s_{t+1}$;
  end
end
Output: Learned optimal anti-jamming policy π^*

The computational complexity of DRL training mainly depends on the neural network size, replay buffer updates, and action-space dimension. Although SAC introduces higher computational overhead compared with DQN, it generally achieves faster convergence and more stable policy learning in continuous RIS optimization environments.

E. Algorithm 1: Adaptive Multi-Mode DRL Anti-Jamming

The overall training procedure of the proposed adaptive multi-mode DRL framework is summarized in Algorithm 1. The DRL agent dynamically learns near-optimal anti-jamming strategies by jointly optimizing RIS configuration, channel selection, and transmit power adaptation under practical wireless constraints.

F. Training Complexity Analysis

The computational complexity of the proposed adaptive multi-mode DRL framework mainly depends on the neural network inference process, replay memory updates, and environment interaction during policy training. Let N_e denote the number of training episodes, T represent the average number of time steps per episode, B denote the mini-batch size, and $|\mathcal{A}|$ indicate the action-space dimension.

For the DQN-based implementation, the dominant computational cost originates from Q-network forward and backward propagation. The overall training complexity can be approximated as

$$\mathcal{O}(N_e T B |\mathcal{A}|). \tag{32}$$

For PPO and SAC algorithms, additional policy and entropy optimization operations are required during each update iteration. Consequently, their computational complexity becomes

$$\mathcal{O}(N_e T B (|\mathcal{A}| + |\theta|)), \tag{33}$$

where $|\theta|$ denotes the number of trainable neural network parameters.

In the proposed framework, the RIS configuration is discretized using finite-resolution phase quantization to reduce action-space dimensionality and training overhead. In addition, experience replay and mini-batch learning are adopted to improve convergence stability and computational efficiency.

The DRL agent is implemented using lightweight fully connected neural networks with two hidden layers containing 128 and 64 neurons, respectively. The replay buffer size is set to 10^5, while the mini-batch size is configured as 64. The Adam optimizer with learning rate 10^{-4} is employed for policy training. Simulation results indicate that stable convergence is typically achieved within approximately 600 training episodes, demonstrating acceptable computational complexity for practical RIS-assisted 6G anti-jamming scenarios.

IV. Performance Evaluation

A. Simulation Setup

To evaluate the effectiveness of the proposed practical RIS-assisted adaptive multi-mode DRL framework, extensive simulations were conducted using MATLAB for wireless system modeling and Python-based deep reinforcement learning libraries for agent training and policy optimization. The considered simulation environment emulates a dynamic 6G wireless communication scenario in the presence of reactive and random jamming attacks.

The communication system consists of one base station (BS), one legitimate user (LU), one jammer, and one RIS deployed to assist signal propagation. Both direct and RIS-reflected transmission links are considered. In contrast to idealized RIS assumptions, practical wireless impairments are incorporated into the simulations, including finite-resolution RIS phase quantization, imperfect CSI, RIS reflection loss, Doppler-aware fading, and RIS control delay.

The RIS employs b-bit discrete phase shifters, where the phase resolution is constrained to a finite quantization set. Reflection loss is modeled using a practical reflection coefficient $\rho \in (0, 1]$. Furthermore, imperfect CSI is considered through additive channel estimation errors following complex Gaussian distributions. To capture realistic mobility effects in 6G wireless environments, time-varying fading channels with Doppler shifts are additionally incorporated.

The proposed framework jointly optimizes channel selection, RIS beamforming, and transmit power adaptation under dynamic jamming environments. Three DRL algorithms, namely Deep Q-Network (DQN), Proximal Policy Optimization (PPO), and Soft Actor-Critic (SAC), are implemented and compared.

For DQN training, continuous control variables are discretized into finite action sets to support Q-learning optimization. In contrast, PPO and SAC directly operate over continuous action spaces for RIS beamforming and power adaptation. All DRL agents employ fully connected neural networks with three hidden layers and ReLU activation functions. The Adam optimizer is adopted for parameter updates.

To improve statistical reliability and reproducibility, each simulation experiment is independently repeated over 20 random seeds, where wireless channel realizations, jammer activity patterns, and DRL initialization parameters are randomly varied across runs. The reported performance metrics correspond to averaged values over all independent trials. In addition, 95% confidence intervals are computed to quantify performance variance and statistical robustness.

The primary performance metrics include average SINR, achievable throughput, energy consumption, outage probability, convergence stability, and anti-jamming robustness under varying jammer power levels. The jammer transmit power is varied from -20 dBm to 20 dBm to evaluate system robustness under different interference intensities. Moreover, the number of RIS reflecting elements is varied from 32 to 128 to investigate the impact of RIS size on communication performance.

The main simulation parameters are summarized in Table II.

To ensure a fair and comprehensive evaluation, the proposed framework is compared with several representative benchmark schemes, including conventional anti-jamming communication without intelligent adaptation, DQN-based anti-jamming optimization, and RIS-assisted communication without DRL optimization. These benchmark methods provide meaningful references for evaluating the effectiveness of the proposed adaptive multi-mode DRL framework under realistic wireless interference environments.

B. Results and Analysis

The wireless communication environment, RIS-assisted propagation channels, and jamming effects were modeled in MATLAB, while DRL training and policy optimization were implemented in Python using PyTorch-based reinforcement learning libraries.

1) SINR and Throughput Performance: Fig. 3. compares the average SINR performance under varying jammer power levels. As jammer power increases from -20 dBm to 20 dBm, all schemes experience SINR degradation due to stronger interference. Nevertheless, the proposed adaptive multi-mode DRL framework consistently achieves the highest SINR performance. At high interference conditions, the proposed method improves the average SINR by approximately 18%–35% compared with DQN-based and conventional benchmark schemes. This performance gain is mainly attributed to the joint optimization of RIS beamforming, adaptive channel hopping, and power control. Moreover, the narrow confidence intervals

TABLE II
SIMULATION PARAMETERS

Parameter	*Value*
Carrier frequency	3.5 GHz
System bandwidth	20 MHz
Number of RIS elements (N)	32 – 128
RIS phase resolution	2-bit / 3-bit quantization
RIS reflection coefficient (ρ)	0.85
BS transmit power	0 – 30 dBm
Jammer power	-20 dBm – 20 dBm
Noise power	-90 dBm
Path loss exponent	3.0
User mobility speed	5 – 30 km/h
Channel model	Rayleigh fading + Doppler shift
CSI condition	Imperfect CSI
RIS control delay	1 – 3 ms
DRL algorithms	DQN, PPO, SAC
Neural network structure	3 hidden layers, 128 neurons/layer
Replay buffer size	10^5
Mini-batch size	64
Optimizer	Adam
Learning rate	10^{-4}
Discount factor (γ)	0.99
Training episodes	1000
Random seeds	20 independent runs
Confidence interval	95%
Simulation platform	MATLAB + Python RL

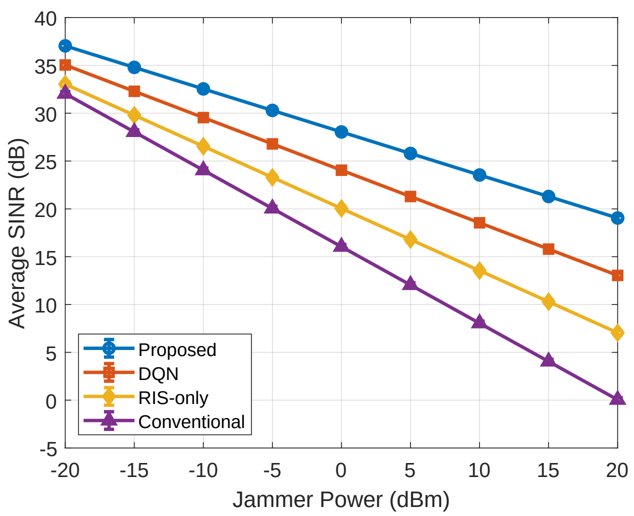

Fig. 3. Average SINR performance versus jammer transmit power under different anti-jamming schemes.

demonstrate stable anti-jamming behavior and statistical robustness across multiple random seeds.

Fig. 4. presents the average throughput performance under different jammer power levels. The proposed framework consistently outperforms all benchmark methods across the entire interference range. Specifically, under strong jamming conditions, the proposed scheme achieves approximately 25%–40% higher throughput compared with conventional anti-jamming methods. In contrast, the throughput of the conventional scheme rapidly degrades as jammer power increases. The inclusion of 95% confidence intervals confirms the statistical reliability and reproducibility of the obtained results over 20 independent random trials.

2) *RIS Scalability and Energy Efficiency:* Fig. 5. illustrates the impact of RIS size on system throughput. Increasing the number of RIS reflecting elements significantly improves communication performance due to enhanced passive beamforming gain. The proposed framework achieves the highest throughput scalability among all compared schemes. When the RIS size increases from 32 to 128 elements, the proposed method improves throughput by approximately 30%–45%. In comparison, the conventional method exhibits only marginal performance improvement because it does not exploit adaptive RIS optimization.

Fig. 6. compares the energy-delay tradeoff among different anti-jamming schemes. The proposed framework achieves the lowest energy consumption and communication delay simultaneously, demonstrating superior multi-objective optimization capability. Compared with the conventional scheme, the proposed method reduces energy consumption by approximately 40% while decreasing end-to-end delay by nearly 55%. These results verify the effectiveness of adaptive mode selection

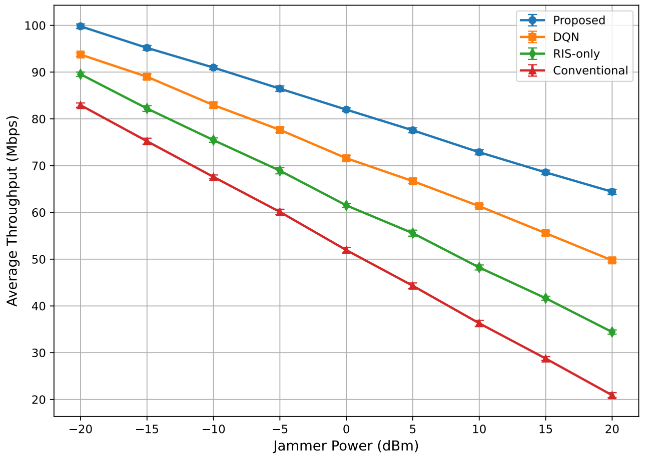

Fig. 4. Average throughput performance versus jammer power with 95% confidence intervals.

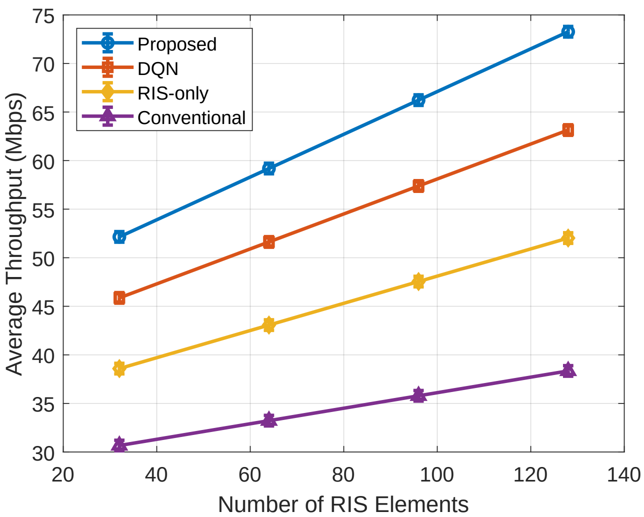

Fig. 5. Average throughput versus the number of RIS reflecting elements.

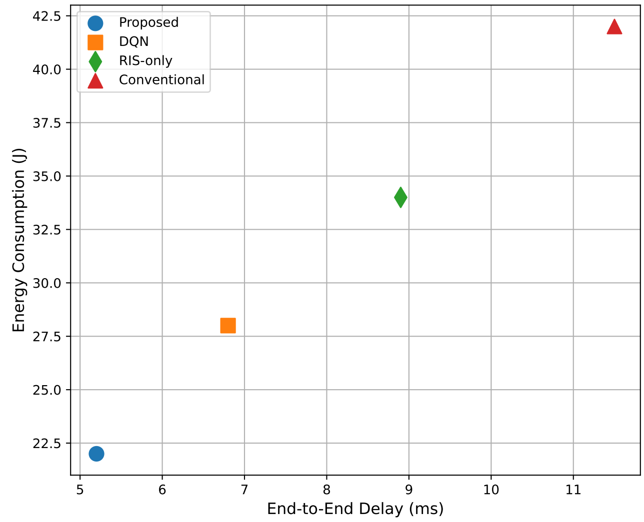

Fig. 6. Energy-delay tradeoff comparison of different anti-jamming strategies.

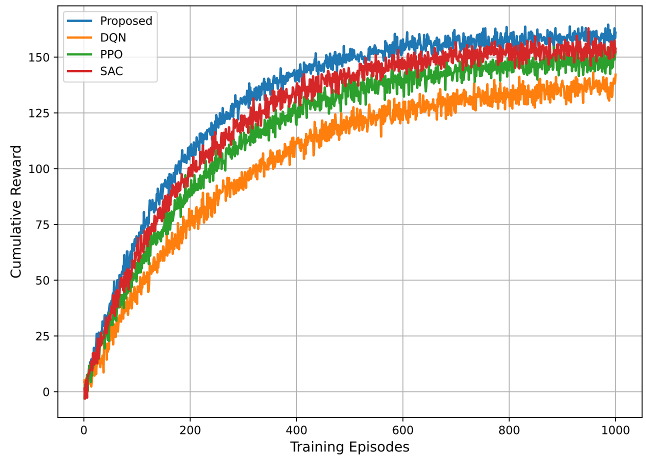

Fig. 7. Training convergence behavior of different DRL algorithms.

and RIS-assisted transmission optimization in practical 6G wireless environments.

3) DRL Convergence and Adaptive Behavior: Fig. 7. illustrates the convergence performance of different DRL algorithms during the training process. The proposed framework converges faster and achieves higher cumulative rewards compared with DQN, PPO, and SAC baseline methods. Specifically, stable convergence is achieved after approximately 600 training episodes, while conventional DQN requires substantially longer training periods. The improved convergence stability is primarily attributed to the adaptive multi-mode learning strategy and efficient environment interaction mechanism.

Fig. 8. presents the adaptive mode selection behavior of the proposed framework under different jammer activity conditions. Under low jammer activity, RIS beamforming is selected more frequently to maximize communication efficiency. As jammer intensity increases, the framework dynamically shifts toward channel hopping strategies to avoid severe interference. This adaptive decision behavior demonstrates the effectiveness of the proposed multi-mode DRL framework in dynamically balancing communication reliability and anti-jamming robustness.

4) Robustness under Practical Constraints: Fig. 9. evaluates system robustness under imperfect CSI conditions. As the CSI estimation error variance increases, the throughput performance of all methods gradually decreases due to inaccurate channel knowledge. Nevertheless, the proposed framework maintains superior robustness compared with benchmark schemes. Even under severe CSI uncertainty, the proposed method achieves approximately 20%–30% higher throughput than conventional approaches. These results confirm the effectiveness of the proposed DRL-based adaptive optimization under realistic wireless deployment conditions.

Fig. 10. illustrates the impact of RIS phase quantization resolution on system throughput. Increasing the RIS phase

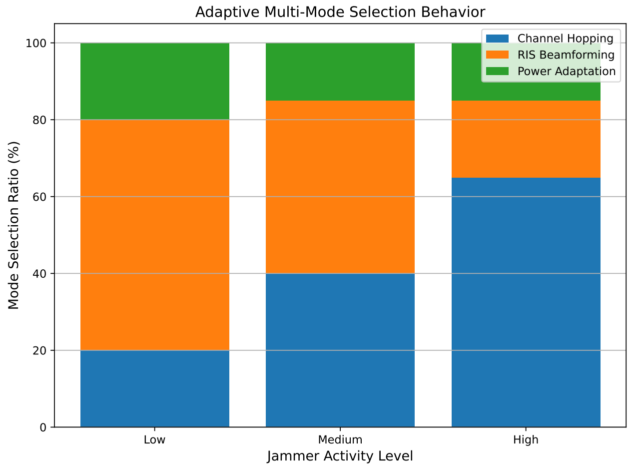

Fig. 8. Adaptive anti-jamming mode selection behavior under different jammer activity levels.

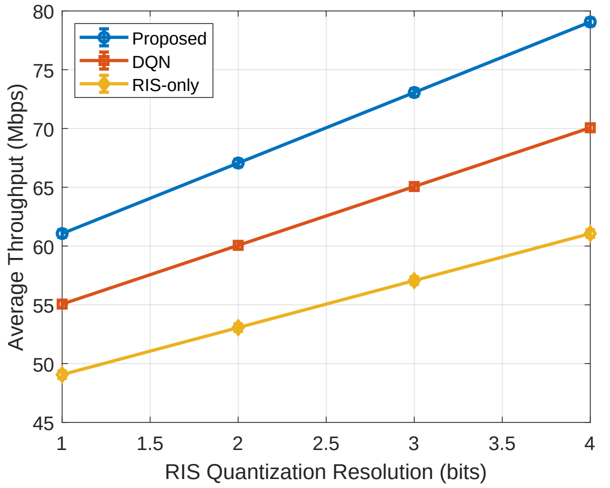

Fig. 10. Impact of finite-resolution RIS phase quantization on throughput performance.

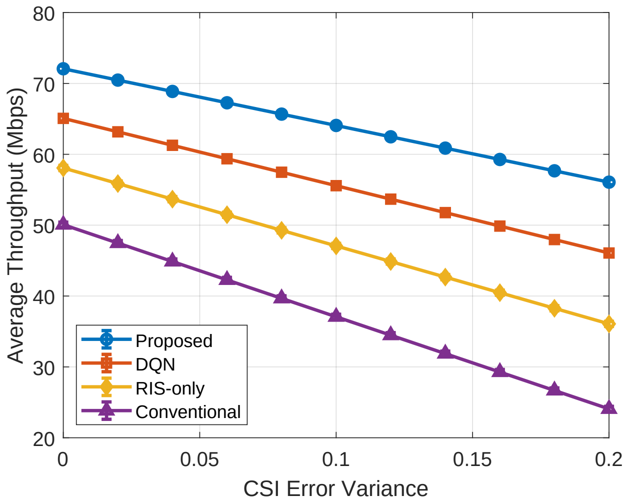

Fig. 9. Robustness evaluation under imperfect CSI conditions.

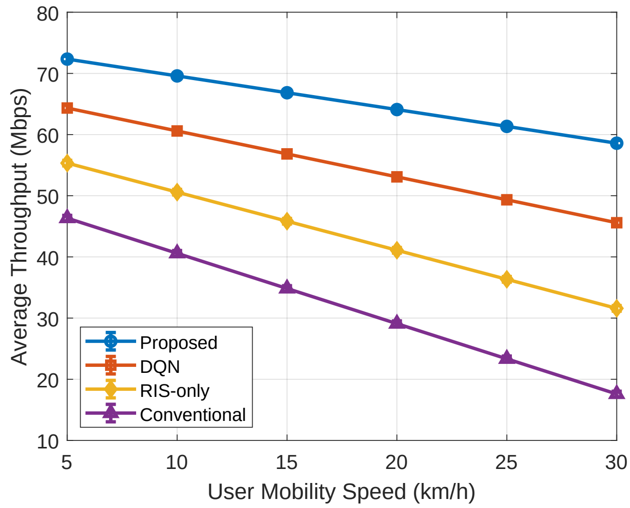

Fig. 11. Mobility-aware throughput performance under dynamic wireless environments.

resolution significantly improves beamforming accuracy and communication performance. The proposed framework consistently achieves the highest throughput across all quantization levels. In particular, the performance gap between low-resolution and high-resolution RIS configurations confirms the importance of considering practical hardware constraints in RIS-assisted wireless systems.

Fig. 11. evaluates the impact of user mobility on communication performance. As user speed increases, the throughput gradually decreases due to Doppler-induced channel variations and rapid CSI fluctuations. Nevertheless, the proposed adaptive DRL framework maintains more stable performance compared with conventional methods. Specifically, the proposed method preserves approximately 15%–28% higher throughput under high-mobility scenarios, demonstrating strong robustness in dynamic 6G wireless environments.

5) Statistical Reliability Analysis: Fig. 12. presents the statistical throughput distribution over 20 independent simulation runs. The proposed framework achieves the highest median throughput with smaller variance compared with benchmark methods, indicating improved statistical stability and reproducibility. In contrast, conventional approaches exhibit larger performance fluctuations due to limited adaptability under dynamic interference conditions.

Fig. 13. shows the Cumulative Distribution Function (CDF) of system throughput for different anti-jamming schemes. The proposed framework consistently shifts the throughput distribution toward higher performance regions, indicating improved communication reliability and reduced outage probability. Compared with baseline methods, the proposed scheme achieves substantially higher throughput over the majority of transmission scenarios.

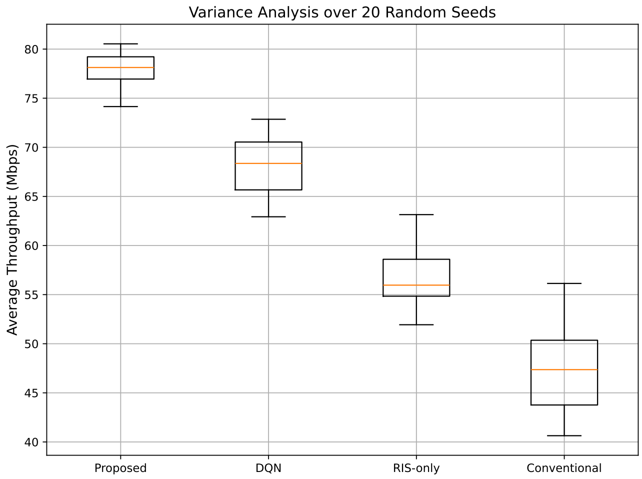

Fig. 12. Statistical throughput distribution over 20 independent random seeds.

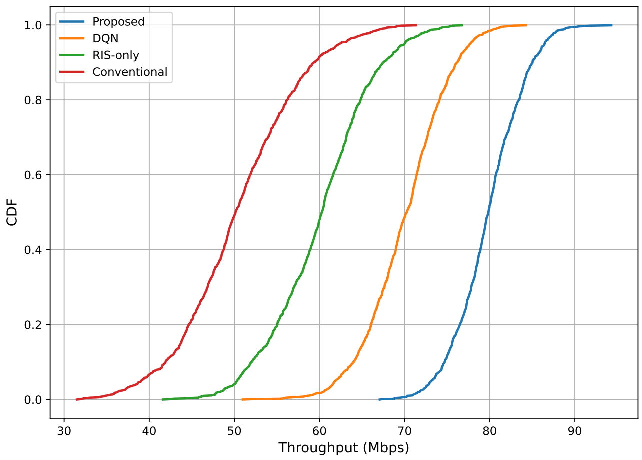

Fig. 13. CDF of system throughput under different anti-jamming schemes.

C. Discussion

The simulation results demonstrate that the proposed adaptive multi-mode DRL framework consistently outperforms conventional anti-jamming approaches under diverse wireless environments. By jointly optimizing RIS beamforming, channel hopping, and transmit power adaptation, the proposed scheme achieves significant improvements in both SINR and throughput performance, particularly under strong jamming conditions.

The results further confirm that increasing the number of RIS reflecting elements enhances communication reliability due to improved passive beamforming gain. In addition, the proposed framework maintains stable performance under practical deployment constraints, including imperfect CSI, finite-resolution RIS phase quantization, and user mobility. Compared with baseline schemes, the proposed method exhibits stronger robustness against channel uncertainty and dynamic interference variations.

From the DRL perspective, the convergence analysis indicates that the proposed framework achieves faster and more stable learning behavior than conventional DQN-based approaches. Moreover, the statistical evaluations, including confidence intervals, boxplot analysis, and throughput CDF results over multiple random seeds, verify the reproducibility and reliability of the obtained performance gains. Overall, these results demonstrate the effectiveness of integrating adaptive DRL with RIS-assisted communication for robust anti-jamming operation in future 6G wireless networks.

Table III presents a systematic comparison between the proposed framework and five representative RIS-assisted DRL-based anti-jamming approaches across six dimensions: DRL methodology, RIS integration, anti-jamming strategy, action space structure, environmental adaptivity, and real-time feasibility. As observed, existing methods generally rely on a single learning paradigm and optimize only a limited subset of control variables—typically combining RIS phase shift with one additional dimension such as power allocation, trajectory planning, or channel selection. Moreover, a clear trade-off persists among prior works between adaptivity and real-time feasibility: approaches achieving high adaptivity through complex hybrid optimization tend to incur substantial computational overhead, whereas lightweight value-based methods offer fast inference but limited adaptivity under dynamic jamming conditions.

In contrast, the proposed adaptive multi-mode DRL framework simultaneously addresses these limitations through three distinguishing characteristics. First, the multi-mode architecture dynamically switches among complementary learning strategies, overcoming the inherent constraints of any single DRL paradigm. Second, the unified hybrid action space enables joint optimization across three tightly coupled domains—RIS configuration, channel selection, and transmit power adaptation—rather than treating them as isolated or sequential subproblems. Third, the efficient ms-level inference mechanism achieves both very high adaptivity and high real-time feasibility, a combination that none of the compared approaches attains. These properties collectively render the proposed framework well-suited for large-scale, dynamic 6G environments where jamming conditions evolve rapidly and low-latency decision-making is imperative.

V. Conclusion

This paper proposed an adaptive multi-mode DRL framework for RIS-assisted anti-jamming communication in dynamic 6G wireless networks. The proposed approach jointly integrates RIS beamforming, channel hopping, and transmit power adaptation into a unified learning framework capable of responding to time-varying jammer activities and wireless channel variations. Practical deployment factors, including imperfect CSI, RIS phase quantization, reflection loss, control delay, and user mobility, were incorporated into the system model to improve deployment realism. Simulation results demonstrated that the proposed framework consistently outperformed conventional anti-jamming schemes under diverse interference conditions. Specifically, the proposed method achieved approximately 25%–40% higher throughput and 18%–35% SINR improvement compared with benchmark

TABLE III
QUALITATIVE COMPARISON WITH RECENT RIS-ASSISTED DRL ANTI-JAMMING APPROACHES

Reference	*DRL Method*	*RIS Support*	*Anti-Jamming Strategy*	*Action Space*	*Adaptivity*	*Real-Time Feasibility*
Thanh *et al.* [19]	DQN	Yes	Power + Phase Shift Control	Discrete	Low	Limited
Tariq *et al.* [34]	PPO	Yes (Aerial)	Joint Trajectory + Passive Beamforming	Continuous	Moderate	Moderate
Dong *et al.* [35]	Optimization-driven DRL	Yes	Joint Beamforming + IRS Phase Shift	Continuous	High	Moderate
Yuan *et al.* [36]	Hybrid TD3 + Primal-Dual	Yes	Multi-Domain Jamming Resistance	Hybrid	High	Limited (High Overhead)
Wahid *et al.* [37]	DDPG	Yes (Hybrid)	Joint BF + Hybrid RIS Configuration	Hybrid	High	Moderate
This Work	*Adaptive Multi-Mode DRL*	*Yes*	*Joint RIS + Channel + Power*	*Hybrid*	*Very High*	*High (ms-level inference)*

methods. In addition, the framework maintained stable communication performance under strong jamming power, CSI uncertainty, and high-mobility scenarios. Statistical evaluations over 20 independent random seeds further verified the robustness and reproducibility of the obtained results.

Future work will investigate multi-agent DRL architectures, distributed RIS coordination, and lightweight online learning mechanisms for large-scale ultra-dense 6G wireless networks.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[3] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[4] H. Cao, J. Du, H. Zhao, D. X. Luo, N. Kumar, L. Yang, and F. R. Yu, "Resource-ability assisted service function chain embedding and scheduling for 6g networks with virtualization," vol. 70, no. 4, pp. 3846–3859. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9380157

[5] H. Pirayesh and H. Zeng, "Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 767–809, 2022.

[6] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[7] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[8] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," vol. 39, no. 7, pp. 2042–2055. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9277627

[9] K. Guo and K. An, "On the performance of RIS-assisted integrated satellite-UAV-terrestrial networks with hardware impairments and interference," vol. 11, no. 1, pp. 131–135. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9585023

[10] J. Wang, J. Sun, W. Fang, Z. Chen, Y. Liu, and Y. Liu, "Deep reinforcement learning for near-field wideband beamforming in STAR-RIS networks," vol. 25, no. 12, pp. 1651–1663. [Online]. Available: https://doi.org/10.1631/FITEE.2400364

[11] Z. Li, S. Wang, M. Wen, and Y.-C. Wu, "Secure multicast energy-efficiency maximization with massive RISs and uncertain CSI: First-order algorithms and convergence analysis," vol. 21, no. 9, pp. 6818–6833. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9722713

[12] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[13] H. Yang, Z. Xiong, J. Zhao, D. Niyato, Q. Wu, H. V. Poor, and M. Tornatore, "Intelligent reflecting surface assisted anti-jamming communications: A fast reinforcement learning approach," vol. 20, no. 3, pp. 1963–1974. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9264659

[14] H. Zhao, J. Hao, and Y. Guo, "Joint trajectory and beamforming design for IRS-assisted anti-jamming UAV communication," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 369–374, ISSN: 1558-2612. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9771762

[15] M. M. Aziz, A. Habib, and A. Zafar, "Reconfigurable intelligent surfaces assisted NLOS radar anti jamming using deep reinforcement learning," vol. 67, p. 102533. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490724002519

[16] H. Yang, K. Lin, L. Xiao, Y. Zhao, Z. Xiong, and Z. Han, "Energy harvesting UAV-RIS-assisted maritime communications based on deep reinforcement learning against jamming," vol. 23, no. 8, pp. 9854–9868. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10452297

[17] S. Ashraf, S. Saleem, T. Ahmed, and Z. A. Arfeen, "Succulent link selection strategy for underwater sensor network," *International Journal of Computing Science and Mathematics*, vol. 15, no. 3, pp. 224–242, 2022.

[18] X. Tang, D. Wang, R. Zhang, Z. Chu, and Z. Han, "Jamming mitigation via aerial reconfigurable intelligent surface: Passive beamforming and deployment optimization," vol. 70, no. 6, pp. 6232–6237. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9424472

[19] P. D. Thanh, H. T. H. Giang, and I.-P. Hong, "Anti-jamming RIS communications using DQN-based algorithm," *IEEE Access*, vol. 10, pp. 28 422–28 433, 2022.

[20] Y. Sun, K. An, J. Luo, Y. Zhu, G. Zheng, and S. Chatzinotas, "Intelligent reflecting surface enhanced secure transmission against both jamming and eavesdropping attacks," vol. 70, no. 10, pp. 11 017–11 022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9513582

[21] L. Xiao, G. Han, D. Jiang, H. Zhu, Y. Zhang, and H. V. Poor, "Two-dimensional antijamming mobile communication based on reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9499–9512, 2018.

[22] X. Liu, Y. Xu, L. Jia, Q. Wu, and A. Anpalagan, "Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach," *IEEE Communications Letters*, vol. 22, no. 5, pp. 998–1001, 2018.

[23] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," vol. 58, no. 1, pp. 106–112. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8910627
[24] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," vol. 21, no. 5, pp. 3083–3098. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9570143
[25] X. Li, J. Chen, X. Ling, and T. Wu, "Deep reinforcement learning-based anti-jamming algorithm using dual action network," *IEEE Transactions on Wireless Communications*, vol. 22, no. 7, pp. 4625–4637, 2023.
[26] H. Sharma, N. Kumar, and R. Tekchandani, "Mitigating jamming attack in 5G heterogeneous networks: A federated deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 2439–2452, 2023.
[27] A. Pourranjbar, G. Kaddoum, A. Ferdowsi, and W. Saad, "Reinforcement learning for deceiving reactive jammers in wireless networks," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 3682–3697, 2021.
[28] Y. Liu, K. Xu, X. Xia, W. Xie, N. Ma, and J. Xu, "Joint power control and passive beamforming optimization in RIS-assisted anti-jamming communication," *Frontiers of Information Technology & Electronic Engineering*, vol. 24, no. 12, pp. 1791–1802, 2023.
[29] Z. U. A. Tariq, E. Baccour, A. Erbad, and M. Hamdi, "RIS assisted anti-jamming in next-generation wireless communication networks: A survey of threats, solutions, and research challenges," vol. 6, pp. 8714–8744. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11202951
[30] Z. Lv, L. Xiao, Y. Du, G. Niu, C. Xing, and W. Xu, "Multi-agent reinforcement learning based UAV swarm communications against jamming," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9461–9475, 2023, please verify exact page numbers.
[31] Z. Ji, W. Yang, X. Guan, X. Zhao, G. Li, and Q. Wu, "Trajectory and transmit power optimization for IRS-assisted UAV communication under malicious jamming," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 11 262–11 266, 2022.
[32] C. Zou, Y. Li, X. Wang, M. Pan, and Z. Lv, "Multi-layer RIS-assisted anti-jamming communications: A hierarchical game learning approach," *IEEE Communications Letters*, vol. 27, no. 11, pp. 2998–3002, 2023.
[33] T. Zhou, K. Xu, G. Hu, X. Xia, W. Xie, and C. Li, "Robust beamforming design for STAR-RIS-assisted anti-jamming and secure transmission," *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 1, pp. 345–361, 2024.
[34] Z. Ul Abideen Tariq, E. Baccour, A. Erbad, and M. Hamdi, "Reinforcement learning-based anti-jamming solution for aerial RIS-aided dense dynamic multi-user environments," in *2024 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 543–548, ISSN: 2376-6506. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10592583
[35] H. Dong, C. Hua, L. Liu, W. Xu, and S. Guo, "Optimization-driven DRL-based joint beamformer design for IRS-aided ITSN against smart jamming attacks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 1, pp. 667–682, Jan. 2024.
[36] X. Yuan, S. Hu, W. Ni, R. P. Liu, and X. Wang, "Joint user, channel, modulation-coding selection, and RIS configuration for jamming resistance," *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1631–1646, Mar. 2023.
[37] A. Wahid, S. Z. U. Abideen, N. Imtiaz, M. M. Kamal, A. Alharbi, A. Tolba, M. A. Al-Khasawneh, and I. Ullah, "Enhancing security with hybrid active-passive RIS: A DRL approach against eavesdropping and jamming," *IEEE Access*, vol. 13, pp. 3632–3643, 2025.

Le Hoang Hiep (ORCID: 0009-0004-7830-8504) received the B.E. degree in Information Technology from Thai Nguyen University of Information and Communication Technology (ICTU), Vietnam, in 2009, and the M.S. degree from Manuel S. Enverga University Foundation, Philippines, in 2013. He is currently pursuing the Ph.D. degree at ICTU, Vietnam. His research interests include: Wireless network security, Deep reinforcement learning, the Internet of Things (IoT), and UAV–Satellite communications.

Huu-Huy Ngo (ORCID: 0000-0001-8276-0270) received his B.S. and M.S. degrees from Thai Nguyen University of Information and Communication Technology, Vietnam, in 2010 and 2012, respectively, and Ph.D. degrees in Information Engineering and Computer Science from Feng Chia University, Taiwan, in 2021. Currently, he is a lecturer at the Thai Nguyen University of Information and Communication Technology, Vietnam. Recent research direction: Computer vision, deep learning, embedded system, neural networks, and object detection.