Asian Journal of Electrical and Electronic Engineering

VOLUME 6 ISSUE 1 JUNE 2026

eISSN 2785-8189

https://alambiblio.com/ojs/index.php/ajoeee

ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING

AIMS & SCOPE OF THE ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING

The **Asian Journal of Electrical and Electronic Engineering** (**AJoEEE**), published biannually (March and September), is a peer-reviewed open-access journal of the **AlamBiblio Press**.

The Asian Journal of Electrical and Electronic Engineering publishes original research findings as regular papers and review papers (by invitation). The Journal provides a platform for Engineers, Researchers, Academicians, and Practitioners who are highly motivated to contribute to the Electrical and Electronics Engineering disciplines. It also welcomes contributions that address the developing world's specific challenges and address science and technology issues from a multidisciplinary perspective.

REFEREES' NETWORK

All papers submitted to AJoEEE Journal will be subjected to a rigorous reviewing process through a worldwide network of specialized and competent referees. Each accepted paper should have at least two positive referees' assessments.

SUBMISSION OF A MANUSCRIPT

A manuscript should be submitted online to the Asian Journal of Electrical and Electronic Engineering (AJoEEE) website https://journals.alambiblio.com/ojs/index.php/ajoeee/. Further correspondence on the status of the paper could be done through the journal's website

COPYRIGHT NOTICE

Consent to publish: The Author(s) agree to publish their articles with AlamBiblio Press.

Declaration: The Author(s) declare that the article has not been published before in any form. It is not concurrently submitted to another publication and does not infringe on anyone's copyright. The author (s) holds the AlamBiblio Press and Editors of the Journal harmless against all copyright claims.

Transfer of copyright: The Author(s) hereby agree to transfer the article's copyright to **AlamBiblio Press**, which shall have the non-exclusive and unlimited right to publish the article in any form, including in electronic media. For articles with more than one author, the corresponding author confirms that he/she is authorized by his/her co-author(s) to grant this transfer of copyright.

The Asian Journal of Electrical and Electronic Engineering follows the open access policy.

All articles published with open access will be immediately and permanently free for everyone to read, download, copy and distribute.

ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING

eISSN 2785-8189

Published by:

AlamBiblio Press,
PV8 platinum Hill
Jalan Melati Utama, 53100 Kuala Lumpur, Malaysia
Phone (+603) 2713 7308

ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING

Volume 6, Issue 1, June 2026

TABLE OF CONTENTS

ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING
Vol. 6 No. 1 2026
E-ISSN: 2785-8189

Hybrid Deep Reinforcement Learning for RIS-Assisted 6G IoT Networks: A Comparative Study of DDPG, TD3, and Adaptive Hybrid Policies

Ezdihar Osman Taj Almowla Mohomad[1*], Khalid Hamid Bilal[2], Zeinab Mahmoud Omer[1], Eltaf Abdalsalam Mohamed[3], and Rania Ali Elkhidir[4]

[1]*University of Bahri Khartoum, Sudan*
[2]*University of Science & Technology, Omdurman*, Sudan

[3]*Blue Nile University, Damazee, Sudan*
[4]*University of Hail, Saudi Arabia*

*Corresponding author: Ezdiharosman22@gmail.com

(Received: 24 February 2026; Accepted: 25 May 2026)

Abstract—Reconfigurable Intelligent Surfaces (RIS) are becoming essential for improving coverage and signal reliability in the upcoming 6G wireless networks. In this research, we present a hybrid deep reinforcement learning (DRL) framework to optimize beamforming in RIS-assisted systems. It employs several policy-selection mechanisms to enhance performance stability and reward consistency under dynamic channel conditions. The extensive experimental evaluation over the last 30 testing episodes used trimmed-mean analysis with 95% confidence intervals.

With an average return of 14.02 ± 0.62, the Hybrid Best-Action method outperformed the conventional DDPG baseline (11.13 ± 0.87) by 25.97%, which is marked by a significant effect size (Cohen's d = 1.42, $p < 0.0001$). Although TD3 achieved a competitive performance (13.34 ± 0.73), the hybrid strategy outperformed it in reward stability and reduced performance variability, which is evidenced by the rolling standard deviation analysis. The results of a one-way ANOVA showed statistically significant differences among all the policies assessed ($F(4,145) = 9.8$, $p < 0.0001$), indicating a considerable overall effect size ($\eta^2 \approx 0.213$).

A post hoc power analysis indicates that the sample size we selected (n = 30 per policy) has high statistical power (>0.99) to detect moderate-to-large performance differences. In RIS-assisted IoT applications, the proposed hybrid framework's reduced reward variability and improved convergence stability are key factors that enhance the reliability of beam alignment and ensure consistent coverage. The findings show that the proposed hybrid DRL method yields statistically significant and practically meaningful improvements over traditional single-policy methods, making it a strong contender for intelligent beamforming optimization in dynamic 6G wireless environments.

***Keywords*:** Reconfigurable Intelligent Surface (RIS), Deep Reinforcement Learning (DRL), DDPG, TD3. Hybrid Learning, Beamforming Optimization,6G Networks, IoT Coverage Enhancement.

1. INTRODUCTION

The continuous development of wireless communication systems has been driven by the rapid growth of data-intensive and latency-sensitive applications [1]. Despite significant advancements in fifth-generation (5G) networks regarding enhanced mobile broadband, ultra-reliable low-latency communications, and extensive machine-type communications, they encounter increasing challenges from emerging use cases such as large-scale Internet of Things (IoT) [2][3][4], autonomous systems, intelligent sensing, and immersive applications. The evolving requirements drive the transition to sixth-generation (6G) wireless networks, which aim to provide

DOI: https://doi.org/10.69955/ajoeee.2026.v6i1.86

extensive coverage, outstanding reliability, and intrinsic intelligence at both the physical and network levels [5-7].

A defining characteristic of 6G is the shift from conventional communication models to wireless systems that are aware of and able to interact with their surroundings [8-11]. In contrast to 5G, which treats the propagation environment as a passive element [12-14], 6G envisions programmable environments that can actively influence electromagnetic wave propagation. Reconfigurable Intelligent Surfaces (RIS) have emerged as a critical enabling technology [15] [16] [17]. By adjusting the phase shifts of various cost-effective reflective elements, RIS can enhance signal coverage, reduce interference, and improve energy efficiency [18] [19] [20]. Achieving these benefits in dense and dynamic IoT deployments is a complex optimization challenge due to the high-dimensional control space, non-convex system behavior, and rapidly changing channel and interference circumstances.

This research aims to develop a comprehensive hybrid deep reinforcement learning framework for RIS-assisted 6G IoT networks, motivated by identified shortcomings. The primary objective is to integrate the complementary benefits of Deep Deterministic Policy Gradient (DDPG) and Twin Delayed DDPG (TD3) into hybrid decision policies, thereby enabling more stable learning, accelerating convergence, and optimizing performance with a focus on coverage in dynamic channel and interference settings. We present numerous simulation results to compare the proposed hybrid architecture with existing solo DRL methods.

2. RELATED WORK

Initial research endeavors predominantly concentrated on modeling and facilitating communications augmented by RIS. A. Taha, M. Alrabeiah, and A. Alkhateeb wrote "Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning" to look into ways to estimate channels that employ compressive sensing and deep learning to cut down on the expense of training for large-scale RIS deployments [21]. In their paper "Reconfigurable Intelligent Surfaces for Wireless Communications: Overview of Hardware Designs, Channel Models, and Estimation Techniques" [22], Jian et al. provide a thorough overview of RIS hardware architectures, channel models, and estimation methods. This work sets the stage for RIS-assisted systems. These attempts are significant, but they don't solve the problem of long-term RIS control or making decisions when there is interference.

Later, to simplify the calculations, supervised deep learning methods were used. Yang, Liu, and Zhang used deep neural networks trained offline to predict optimal RIS phase configurations based on channel state information in their paper "A Deep Learning-Based Modelling of Reconfigurable Intelligent Surface-Assisted Wireless Communications for Phase Shift Configuration" [23]. Their method works well in static or slowly changing channels, but it requires labeled datasets and struggles to adapt to rapidly changing environments. This means it can't be used in real-world 6G IoT situations.

Reinforcement learning (RL) and deep reinforcement learning (DRL) algorithms have been used to address the challenges of RIS optimization. Zhou, Wang, and Li present a deep reinforcement learning framework in their study, "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Cooperative Jamming Model Design," to improve physical-layer security by optimizing RIS phase shifts [24]. ZUA Tariq, E Baccour, Erbad, and M Hamdi also looked into how to make wireless communication networks more resistant to jamming attacks in their study "Reinforcement Learning for Resilient Aerial-IRS-Assisted Wireless Communications Networks in the Presence of Multiple Jammers" [25]. The results demonstrated that DRL can proficiently address continuous control challenges in dynamic settings. Still, they rely on autonomous learning agents that often struggle with training instability, reward unpredictability, and overestimation bias.

In specific contexts, more utilizations of DRL-based RIS have been investigated. In "Reconfigurable Intelligent Surface-Assisted Localization: Technologies, Challenges, and the Road Ahead," MA and Teng et al. [26]. Examined RIS-enabled localization and emphasized the potential of learning-based optimization. Recent studies, such as "Reinforcement Learning-Based Intelligent Reflecting Surface Optimization for Wireless Communications" [27], employ single-agent reinforcement learning frameworks to improve system performance relative to static baselines. Even with these improvements, these methods often lack thorough statistical tests of their convergence and training stability. Recently, many have suggested using hybrid learning methods to improve performance in "Hybrid Reinforcement Learning for STAR-RISs: A Coupled Phase-Shift Model Based Beamformer" [28], J. Chen et al. presented a hybrid reinforcement learning framework designed to enhance

spectral efficiency for STAR-RIS beamforming. Their research primarily emphasizes throughput measurements, neglecting the assessment of learning stability, reward variance, or performance metrics focused on coverage in densely populated IoT contexts.

From our previous discussion, it's clear that current RIS-assisted learning approaches have major problems. Supervised learning methods aren't very flexible, single-agent deep reinforcement learning methods have problems with stability and convergence when dynamic interference is present, and hybrid approaches don't cover as much ground or go as deep in their evaluations. In the ever-changing world of 6G IoT, it is very hard to ensure that learning remains consistent, that reward variability is low, and that coverage optimization is effective.

3. SYSTEM OVERVIEW AND PROBLEM FORMULATION

3.1 System Overview

We consider a downlink RIS-assisted 6G Internet-of-Things (IoT) network consisting of a base station (BS) equipped with Mantennas, a reconfigurable intelligent surface (RIS) comprising Nnearly passive reflecting elements, and multiple single-antenna IoT devices randomly distributed within the coverage area. The RIS is deployed to enhance wireless propagation by intelligently adjusting the phase shifts of its reflecting elements. Due to blockages, severe path loss, and dense device deployment, the direct BS–IoT links may be weak or unavailable. To address this issue, the RIS establishes an indirect BS–RIS–IoT link, enabling controllable signal reflection and coverage enhancement without additional transmit power. Each RIS element applies a programmable phase shift to the incident signal, allowing the wireless environment to be dynamically reconfigured.

3.2 Channel and Signal Model

Let $\mathrm{G} \in \mathbb{C}^{N\times M}$denote the channel matrix between the BS and the RIS, $\mathrm{h}_k \in \mathbb{C}^{N\times 1}$the channel vector between the RIS and the k-th IoT device, and $\mathrm{d}_k \in \mathbb{C}^{M\times 1}$ the direct BS–IoT channel. The RIS reflection matrix was defined as in [33] using Equation (1).

$$\Phi = \mathrm{diag}\left(e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}\right), \qquad (1)$$

where $\theta_n \in [0,2\pi)$ represents the phase shift applied by the n-th RIS element. The received signal at the k -th IoT device was expressed by Equation (2).

$$y_k = \left(\mathrm{h}_k^H \Phi \mathrm{G} + \mathrm{d}_k^H\right)\mathrm{w}s + n_k, \qquad (2)$$

where w is the BS beamforming vector, s is the transmitted symbol with unit power, and n_k is additive white Gaussian noise with variance σ^2.

3.3 Performance Metric

The signal-to-noise ratio (SNR) at the k-th IoT device is given by

The signal-to-noise ratio (SNR) at the k-th IoT device was calculated using Equation (3).

$$\mathrm{SNR}_k = \frac{|\left(\mathrm{h}_k^H \Phi \mathrm{G} + \mathrm{d}_k^H\right)\mathrm{w}|^2}{\sigma^2}. \qquad (3)$$

Based on the SNR, the achievable rate or coverage-related reward can be derived and used as a performance indicator for RIS optimization.

3.4 Problem Formulation

The objective of the RIS-assisted system is to determine the optimal RIS phase-shift configuration that maximizes the long-term communication performance under dynamic channel conditions. This optimization problem can be formulated as [29]

$$\max_{\Phi} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r_t\right], \qquad (4)$$

subject to

$$\theta_n \in [0,2\pi), \forall n = 1, \dots, N \ , \qquad (5),$$

where r_tIs the reward at the time step t, $\gamma \in (0,1)$is the discount factor, and T is the episode length. Due to the high dimensionality, nonlinearity, and time-varying nature of the wireless environment, solving Equation (4) using conventional optimization techniques is computationally prohibitive. Therefore, the problem was naturally modeled as a Markov Decision Process (MDP), hence, suitable for deep reinforcement learning-based solutions.

3.5 Markov Decision Process Formulation

The tuple$(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P})$ defines the RIS optimization problem as an MDP, where [30]

- State (S): comprises channel-related data, including received SNR, effective channel gains, and prior RIS setups.

The continuous RIS phase-shift adjustments $\theta = [\theta_1, \dots, \theta_N]$.

- Action ($\mathcal{A}$): consists of continuous control variables corresponding to the RIS phase-shift adjustments. At each decision step t, the action is defined as

$$\mathrm{a}_t = \theta_t = [\theta_{1,t}, \theta_{2,t}, \dots, \theta_{N,t}], \qquad (6)$$

where $\theta_{n,t} \in [0,2\pi)$denotes the phase shift applied by the n-th RIS element.

- Reward (R): is intended to represent performance relating to throughput or coverage.
- The stochastic evolution of wireless channels over time is represented by transition (p).
  To learn the best RIS configurations, this formulation enables the implementation of continuous-control DRL algorithms, such as DDPG, TD3, and their hybrids.

4. DRL-BASED JOINT DESIGN UNDER INTERFERENCE DYNAMIC ENVIRONMENTS

4.1 Motivation and Design Rationale

Wireless links in RIS-assisted 6G IoT networks are naturally vulnerable to significant channel fluctuations, dynamic interference, and unanticipated environmental changes due to elevated device density and mobility. These effects make coverage and signal quality much worse, making them look like general jamming-like problems. So, it's very important to develop robust RIS control rules to address these issues. Setting up RIS with traditional optimization methods isn't very flexible when interference conditions change. They also usually assume that settings don't change very often. This encourages the use of deep reinforcement learning (DRL) in situations where you don't know what's going to happen because it lets you change RIS phase shifts and transmission methods in a way that is both adaptive and doesn't require a model.

4.2 Joint DRL-Based RIS Control Framework

The suggested joint design aims to optimize RIS phase shifts to achieve optimal coverage performance in environments where interference varies over time. The DRL agent takes the same control action at every decision point, regardless of the RIS phase shift. It does this by examining the current state of the network, including the quality of the received signal, effective channel conditions, and previous RIS settings. The DRL framework automatically mitigates interference and channel degradation without directly modeling them. It does this by continually adjusting the RIS setup to meet the area's needs. This is not the same as methods for optimizing that don't change.

4.3 Continuous-Control DRL Algorithms

Two sophisticated continuous-control deep reinforcement learning methods are used in this work to address the high-dimensional and persistent aspects of RIS control:

Deep Deterministic Policy Gradient (DDPG): DDPG uses an actor–critic architecture to create deterministic policies in dynamic action spaces. When changes occur quickly, DDPG may struggle with overestimation bias and training instability, but it performs well for control tasks with multiple dimensions. Target policy smoothing, delayed policy updates, and twin critics make the Twin Delayed Deep Deterministic Policy Gradient (TD3) better than DDPG. Training is much more stable and less likely to anticipate an excessively high score because of these characteristics.

The fundamental guidelines for optimizing RIS in scenarios when interference is anticipated are these algorithms.

4.4 Hybrid DRL-Based Joint Design

This paper introduces hybrid decision-making techniques that dynamically combine the benefits of DDPG and TD3 to overcome the shortcomings of independent DRL algorithms. Three kinds of hybrid policies, namely, fixed, best action, and dynamic, were investigated:

- Fixed-α Policy:

A mix of DDPG and TD3 with a fixed- α parameter that makes both the policies do the same thing.

- Best-Action Policy:

A combination of DDPG and TD3 that picks the best course of action based on the highest estimated immediate reward at each point where a decision needs to be made.

- Dynamic-α(t) Policy:

This is a flexible hybrid policy that adjusts the mixing parameter α(t) over time based on learning stability and reward trends. This allows for dynamic policy supremacy. Because they employ multiple DRL strategies, these hybrid systems are less susceptible to channel changes or interference.

5. HYBRID DRL-BASED JOINT DESIGN ALGORITHM

5.1 Overview

This section presents the proposed hybrid deep reinforcement learning (DRL) framework for RIS-assisted 6G IoT networks characterized by variable interference. The framework uses both Deep Deterministic Policy Gradient (DDPG) and Twin Delayed Deep Deterministic Policy Gradient (TD3) to get the best results when optimizing the continuous RIS phase shift.

The agent uses both DDPG and TD3 policies to decide what to do at each decision point, based on the environment's state. A hybrid decision module then uses rules to select policies that are either already set or can be adjusted, thereby determining the final RIS configuration.

5.2 Standalone DRL Policies

1) DDPG-Based RIS Optimization

The DDPG agent has two parts: a critic network that evaluates state-action pairs and an actor network that produces continuous RIS phase-shift actions. DDPG works well for high-dimensional continuous control problems, but its performance may suffer in very dynamic channel and interference situations.

2) TD3-Based RIS Optimization

The TD3 agent builds on the DDPG architecture by using twin critic networks, delayed actor updates, and target policy smoothing. These processes make TD3 stronger in situations where interference is likely by reducing overestimation bias and stabilizing learning.

5.3 Hybrid Decision Policies

Three hybrid decision-making techniques are used to address the problems with standalone DRL policies.

- The Fixed-α Hybrid Policy

:The last RIS action is the weighted sum of the DDPG and TD3 actions

$$\mathrm{a}_t = \alpha \mathrm{a}_t^{\mathrm{DDPG}} + (1-\alpha)\mathrm{a}_t^{\mathrm{TD3}}, \quad (7)$$

.where $\alpha \in [0,1$ is the weight that doesn't change

- The Best Action Hybrid Plan

At every time step, both DDPG and TD3 suggest possible actions. The choice with the highest expected return is :made

$$\mathrm{a}_t = \arg \max_{\mathrm{a} \in \{\mathrm{a}_t^{\mathrm{DDPG}}, \mathrm{a}_t^{\mathrm{TD3}}\}} Q(s_t, \mathrm{a}), \quad (8)$$

.where $Q(\cdot)$shows the rating from the critic

- A policy that combines dynamic-α(t) with other policies

The Dynamic -$\alpha(t)$ policy modifies the mixing parameter according to reward statistics and training :advancement

$$\mathrm{a}_t = \alpha(t)\mathrm{a}_t^{\mathrm{DDPG}} + \big(1-\alpha(t)\big)\mathrm{a}_t^{\mathrm{TD3}}, \quad (9)$$

where $\alpha(t)$ chooses the strategy that works better and is more stable at each step of training.

5.4 Algorithm Description

Algorithm 1 summarizes the proposed hybrid deep reinforcement learning (DRL) optimization framework for RIS-assisted beamforming. The proposed approach combines two continuous-control DRL agents, namely DDPG and TD3, within a unified hybrid decision-making architecture. During each interaction step, both agents independently generate candidate actions based on the observed state of the wireless environment. The hybrid policy module then selects the final RIS configuration action according to the adopted hybrid selection strategy. The generated experiences are stored in replay buffers and used to update the actor–critic networks iteratively throughout the training process.

Figure 1 depicts the general flow of the proposed hybrid DRL framework. The first step in the process is environment initialisation and state observation. In this step, the wireless channel conditions and RIS-related parameters are collected from the communication environment. For RIS phase optimisation, the continuous control actions are separately output by the DDPG and TD3 agents based on the observed state. The generated actions are processed by the Hybrid Decision Module, which leverages the strengths of both DRL policies to select the most suitable RIS configuration strategy. Then, the optimal RIS phase shifts are applied in the wireless environment to enhance signal propagation and communication quality. The environment provides a reward signal and the next state observed after applying the RIS configuration, indicating the communication performance achieved in terms of SINR, coverage quality and system throughput. All collected transition samples are stored in the replay buffer and then used during the policy update stage to iteratively improve learning. This closed-loop interaction is sustained through the training process until convergence is reached.

6. SIMULATION SETUP

6.1 Simulation Environment and Dataset Description

The experimental assessment of the proposed hybrid DRL framework was performed in a simulated RIS-assisted 6G IoT scenario. The dataset used in this work was sourced from Kaggle [34], a platform that offers

publicly available datasets commonly used for benchmarking and reproducible research. The chosen dataset simulates a wireless communication scenario relevant to RIS-enabled 6G IoT networks, including performance metrics such as throughput, latency, energy consumption, and channel-related parameters across diverse channel and interference conditions. To provide an equitable and uniform comparison across all learning policies, data were collected from various simulation episodes with distinct channel realizations and interference levels. The dataset was subsequently processed and saved in NPZ format (R5_NewData_BestHybrid_Results.npz) to enable efficient loading and smooth integration with DRL algorithms, including DDPG, TD3, and the proposed hybrid policies.

The dataset, although simulation-based and derived from a publicly accessible benchmark, encompasses various channel realizations, interference levels, and performance metrics indicative of RIS-assisted IoT contexts. The main aim of this study is to conduct comparative policy evaluation in dynamic settings rather than focus on physical-layer channel modeling. The suggested hybrid framework is architecture-independent and can be easily integrated with physics-based channel models, such as ray-tracing or Deep MIMO datasets, in subsequent implementations.

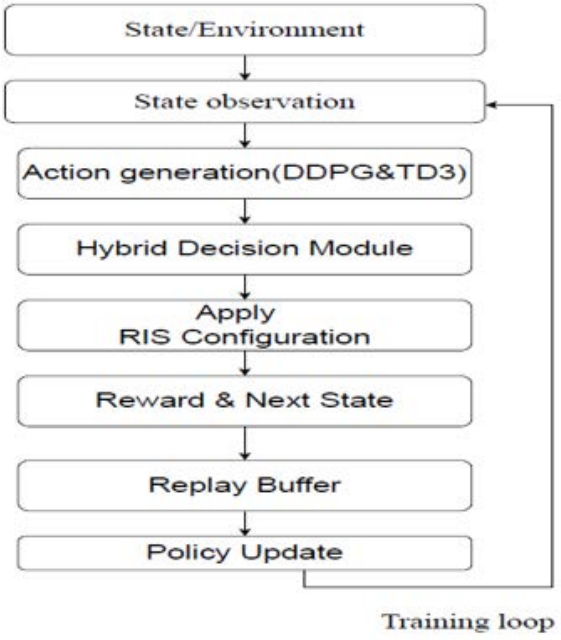

Fig. 1: Workflow of the Proposed Hybrid TD3-DDPG Framework for RIS-Assisted Beamforming Optimization.

6.2 Data Preprocessing and Enhancement

Before training the DRL models, various preprocessing and augmentation techniques were employed to enhance learning stability and mitigate data bias. Initially, partial records and non-numeric entries were discarded to ensure the dataset's uniformity. A feature selection strategy was employed to retain only the metrics pertinent to coverage performance, interference, and transmission efficiency. Third, we standardized all numerical features to a uniform range to prevent high-magnitude variables from dominating and to facilitate faster neural network convergence.

Furthermore, statistical validation was conducted across multiple training iterations to ensure the dataset's capacity to accommodate variations in channel dynamics and interference. The preprocessing approaches ensured that the performance improvements observed during training were primarily attributable to the proposed hybrid learning strategy rather than issues with the input data.

6.3 Feature Description

Table 1 summarizes the key features extracted from the dataset and utilized during the simulation and training process.

Table 1. Dataset Features Used in the Simulation

Feature	Description	Usage in DRL Framework
Throughput (Mbps)	Achievable data rate for IoT devices	Reward maximization
Latency (ms)	End-to-end transmission delay	Penalty term
Energy Consumption (kWh/GB)	Energy efficiency metric	Penalty term
Channel Gain	Composite channel including RIS reflection	State representation
Interference Level	Aggregate interference power	State representation
Noise Power	Thermal noise variance	SNR computation
Episode Return	Cumulative reward per episode	Performance evaluation

6.4 Simulation Advantages

The use of simulation-generated data rather than fixed real-world datasets enables controlled, repeatable experiments, which are essential for evaluating adaptive DRL-based optimization strategies in emerging 6G wireless environments. This setup allows systematic performance analysis under diverse conditions while maintaining experimental consistency.

6.5 Simulation Environment

The performance of the proposed hybrid DRL framework is evaluated using a simulated RIS-assisted 6G IoT environment. The considered network consists of a multi-antenna base station, a reconfigurable intelligent surface with N reflecting elements, and multiple single-antenna IoT devices randomly distributed within the coverage area. Wireless channels follow a block-fading model and vary across episodes to capture dynamic propagation and interference conditions.

All DRL agents are trained under identical environmental settings, channel realizations, and random seeds to ensure a fair and unbiased comparison among different learning policies.

6.6 Training Configuration

Both DDPG and TD3 agents use neural networks with fully connected layers. Target networks help stabilize learning, while experience replay buffers add variety to the training samples. Target policy smoothing, delayed policy updates, and two critics are all used by TD3.

The only thing that makes hybrid policies like Fixed-α, Best-Action, and Dynamic-α(t) different is how they make decisions. All of these policies use the same basic DRL networks. This design ensures that any differences in performance are due solely to the hybrid policy approach, not to changes in the network architecture.

6.7 Protocol for Training and Evaluation

The interaction data (episodes) were split into three parts: training, validation, and testing. This was done to ensure the evaluation was fair. Seventy percent of the episodes were used for policy learning (training), fifteen percent for hyperparameter selection (validation) without changing the agent, and the remaining fifteen percent for final reporting (testing). Also, the assessment was conducted using specific random seeds to reduce variance

and improve reliability. We used trimmed-mean statistics and 95% confidence intervals to assess the final performance across K = 30 testing episodes.

In line with standard statistical power analysis recommendations for identifying moderate-to-large effect sizes in comparative learning studies, we chose K = 30 independent testing episodes to ensure our results are statistically reliable. The post-hoc power analysis we carried out has confirmed that the sample size we used provides statistical power over 0.99, which means it can robustly estimate performance and minimize the chance of a Type-II error.

6.8 Baseline Schemes

The proposed framework is compared against the following baseline schemes:

- No-RIS Configuration: RIS elements are disabled or configured with random phase shifts.
- DDPG-Based Optimization: RIS phase shifts are optimized using standalone DDPG.
- TD3-Based Optimization: RIS phase shifts are optimized using standalone TD3.

These baselines enable systematic evaluation of the performance gains achieved by hybrid decision policies.

6.9 Performance Metrics

We use the following metrics to see how strong a system is and how well it learns:

- The Episode-Based Cumulative Reward shows how well coverage works over time.
- Training Stability: how well things fit together and how much the reward changes.
- The statistical performance is based on how well the rewards are spread out over the episodes.
  These measures, when taken together, show that the agent can learn and adapt to new situations.

7. SIMULATION RESULTS

This section provides a comprehensive performance review of the proposed hybrid deep reinforcement learning (DRL) framework for RIS-enabled 6G IoT networks. The evaluation centers on learning behavior, convergence traits, statistical robustness, and a conclusive performance comparison with standalone DDPG and TD3 benchmarks. To ensure fairness, all schemes are trained and tested under the same simulation settings.

7.1 *Learning Behavior and Convergence Analysis*

Figure 2 shows how the episodic Reward-Trajectories changed over the last 30 evaluation episodes of different reinforcement learning methods. The Hybrid Best-Action-Approach consistently yields the highest rewards over many episodes, demonstrating strong peak performance and greater reward stability. The method seems to work better for handling changes in the channel because it has less oscillation and a more stable path. TD3 and Hybrid Dynamic α(t) both learn quickly, but they differ significantly. However, they don't always get as high of returns as the Hybrid-Best-Action-Method. DDPG doesn't work very well, which quickly degrades rewards. This means the system doesn't adapt quickly to environmental changes, and that beamforming optimization doesn't always work. The Hybrid-Best-Action-Approach reduces randomness, helping the RIS determine the best phase shift. This quickly improves signal coverage and keeps the beam alignment stable. As we move into the 6G-IoT era, factors such as channel fading and mobility could affect performance. To get the best results and reliable coverage, it is important to keep things stable. The episodic behavior indicates that the proposed hybrid action-selection mechanism enhances both optimization efficiency and operational resilience. This makes it a great choice for building advanced wireless networks that use RIS.

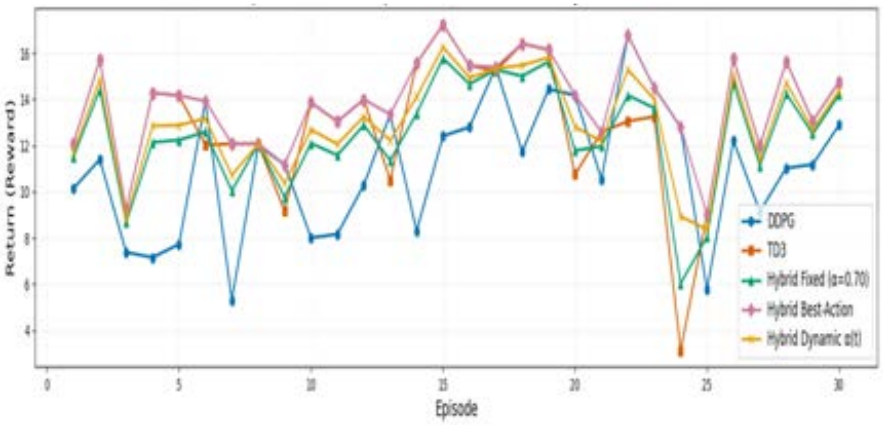

Fig. 2. Episodic reward evolution of evaluated DRL and hybrid policies over the final 30 testing episodes.

7.2 Training Stability Analysis

Figure 3 shows how the rolling standard deviation of returns changes over time. This means that even though the training is the same, the benefits are different in each case. A smaller rolling fluctuation indicates a more stable regulatory environment and a more accurate assessment, especially when channel conditions change.

The Hybrid Best-Action method reduces evaluation variability, making the learning process more stable. TD3 is more susceptible to environmental variations because it is difficult to predict outcomes across sessions. During the first and middle stages of training, DDPG exhibits significant instability. This lengthens the acclimatization period, making it more likely that beamforming changes will fail if not done correctly.

Several studies show that making incentives less random leads to more consistent RIS phase-change choices and more reliable beam alignment. Policy behavior must remain stable in the dynamic 6G IoT environments characterized by channel fading, user mobility, and variable interference. To avoid sudden drops in performance, it's important to ensure that signal coverage is always available. The Hybrid Best-Action approach offers better coverage reliability, improved power distribution, and greater resilience in real RIS-assisted wireless networks, as evidenced by its lower rolling variation.

The stability analysis corroborates the statistical results, showing that the proposed hybrid reinforcement learning framework improves average performance while significantly reducing optimization volatility. This is very important for the practical use of modern smart communication systems to work well.

Fig. 3. Training stability analysis based on rolling standard deviation of episodic returns.

Figure 4 illustrates the distribution of episodic returns for all assessed reinforcement learning policies using a

boxplot. The Hybrid Best-Action strategy demonstrates the highest median return and a reasonably narrow interquartile range, signifying both exceptional central tendency and enhanced performance stability. The distribution is concentrated in the high-reward region, indicating continuous optimization behavior throughout evaluation episodes.

TD3 and Hybrid Dynamic α(t) exhibit similar median performance; however, their broader interquartile ranges indicate marginally greater variability than the Hybrid Best-Action method. Conversely, DDPG has a lower median and higher variance, characterized by significant performance variability and lower-bound outliers, indicating reduced resilience under dynamic channel conditions. The occurrence of sporadic lower outliers in particular policies underscores vulnerability to individual channel realizations, while the Hybrid Best-Action technique ensures more stable high-return results. The boxplot analysis corroborates the statistical findings, demonstrating that the proposed hybrid mechanism yields higher rewards and greater stability in RIS-assisted beamforming optimization contexts.

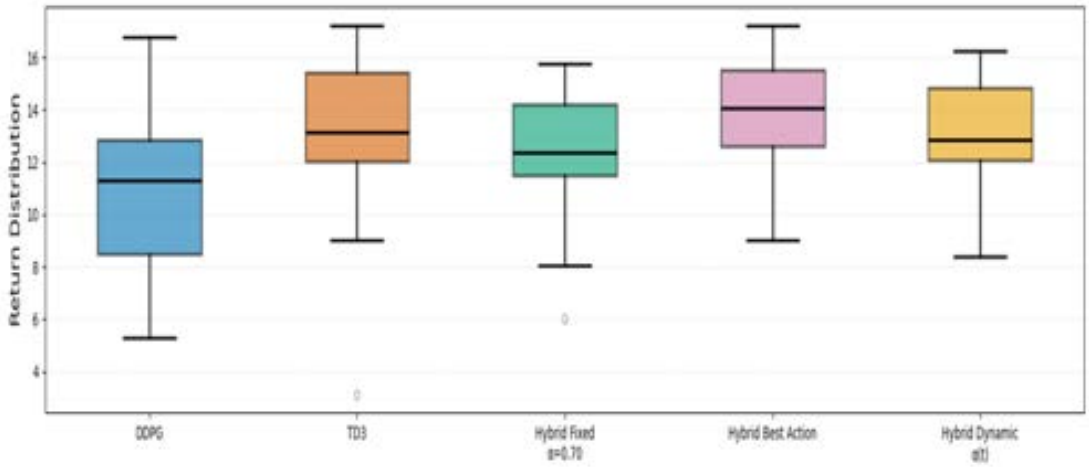

Fig. 4. Distribution of episodic returns for evaluated DRL and hybrid policies (boxplot representation)

7.3 . Statistical Performance and Robustness

Figure 4 shows a box plot of the distribution of returns from each episode. The results show that DDPG has the widest interquartile range and a few low-return outliers. This means that the person doesn't always learn the same way. TD3 limits the range of rewards compared to DDPG, but there is still a lot of difference.

Hybrid policies are better because their return distributions are much narrower. The Best-Action policy differs from the others because it has the highest median return and the smallest spread, indicating greater stability over time. The Dynamic-α(t) policy keeps the distribution small, but it always beats DRL algorithms that don't work together.

7.4 Final Performance Comparison

Figure 5 shows that the Hybrid Best-Action policy has the highest average return among the strategies tested, and its confidence intervals are not very wide. This means that the performance is more consistent. TD3 and Hybrid Dynamic α(t) both perform well, but their average returns are still slightly lower than those of the proposed hybrid approach. DDPG, on the other hand, has the lowest average return and the widest confidence interval. This means that it is less stable and less good at finding the best solution.
The fact that the confidence intervals for Hybrid Best-Action and DDPG don't overlap visually supports the statistically significant improvement indicated by the t-tests and ANOVA. The hybrid method also has tighter error bars, which means that the rewards are more stable from one evaluation episode to the next. This shows that the method still works well even when channel conditions change.
The figure shows that, in general, the proposed Hybrid Best-Action strategy works better and is more stable than standard DRL baselines.

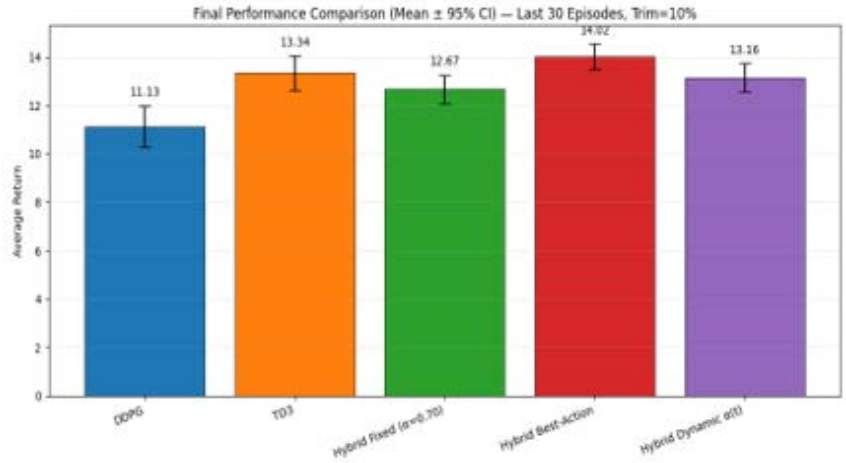

Figure 5. Episode-Based Return Comparison of Standalone and Hybrid DRL Policies in RIS-Assisted 6G IoT Environment.

Table 2. Comparative Performance and Statistical Analysis of Evaluated Policies

Policy	Mean	95% CI	Rel. Improve (%)	Cohen's d	p-value	K
DDPG	**11.13**	**±0.87**	**0%**	**—**	**—**	**30**
TD3	**13.34**	**±0.73**	**+19.86%**	**1.03**	**< 0.001**	**30**
Hybrid Fixed-α	**12.67**	**±0.66**	**+13.84%**	**0.75**	**0.005**	**30**
Hybrid Best-Action	**14.02**	**±0.62**	**+25.97%**	**1.42**	**< 0.0001**	**30**
Hybrid Dynamic-α(t)	**13.16**	**±0.71**	**+18.24%**	**0.95**	**< 0.001**	**30**

According to Table 2, all contemporary DRL and hybrid approaches significantly outperform DDPG. The Hybrid Best-Action technique outperforms the other method both statistically and practically, boasting a remarkable relative improvement of 25.97% and a significant effect size (d=1.42).

7.5 Computational Cost Analysis

To ensure a fair comparison, training time was recorded using Python's time module under the same hardware conditions. With fusion taking place at the action-selection level without adding extra deep network layers, the hybrid policy incurs minimal computational overhead. We assessed the computational efficiency of the suggested DRL algorithms by evaluating the wall-clock training duration under identical experimental settings (Intel i7 CPU, 25,000 timesteps). Table 3 presents an overview of the findings. DDPG required 7.309 minutes, while TD3 required 7.543 minutes, representing a marginal 3.20% increase in training time. The Hybrid (Best-Action) strategy required 14.874 minutes, approximately doubling the computational cost due to the sequential training of both DDPG and TD3 components. Despite the increased cost, the hybrid approach offers improved decision robustness, indicating a trade-off between computational efficiency and performance stability.

Table 3. Comparative Analysis of Training Duration under Identical Experimental Conditions

Algorithm	Training Time (min)	Relative to DDPG
DDPG	7.309	—
TD3	7.543	+3.20%
Hybrid (Best-Action)	14.874	+103.51%

Although the proposed framework was evaluated in a simulation-based environment, it can be extended to practical wireless deployment scenarios using realistic channel generation frameworks such as DeepMIMO and ray-tracing-based propagation models. Integrating the proposed hybrid DRL framework with real-world channel measurements may further improve the robustness and practical applicability of RIS-assisted beamforming optimization in future 6G IoT networks.

8. DISCUSSION

Using the data in Figures 1-5, this section examines how well the hybrid DRL-based RIS optimization framework performs, how stable it is, and how robust it is.

Figure 1 illustrates how the proposed hybrid learning system operates. The system helps TD3 and DDPG agents make decisions by showing them the RIS-assisted wireless environment as a set of choices. The hybrid decision module is responsible for these things. It picks the best control strategy or combines them before setting up RIS. This method promotes adaptive learning and enhances performance in subsequent discoveries without directly replicating interference.

Figures 2 and 4 provide a detailed comparison of episode-based return performance among the independent DDPG, independent TD3, and the proposed hybrid decision policies: Fixed-α, Best-Action, and Dynamic-α(t). The results unequivocally indicate that hybrid strategies consistently yield higher average returns and greater learning stability than DRL-only optimization techniques .The performance trends we discovered correspond with those noted in [31], where the authors examined deep reinforcement learning-based passive beamforming for RIS-assisted wireless networks. The research indicated that TD3 has enhanced stability and superior convergence relative to DDPG, attributable to its twin-critic architecture and reduced overestimation bias. Nonetheless, [31] revealed considerable reward variability under highly dynamic channel and interference settings, highlighting the inadequacies of relying exclusively on a singular learning agent. Figure 2 demonstrates that TD3 achieves higher returns than DDPG, yet it still exhibits variability in performance across episodes .The proposed hybrid decision rules improve TD3-based optimization by dynamically incorporating complementary learning attributes. Figure 2 shows that the boxplot analysis indicates that hybrid strategies display superior median returns and reduced interquartile ranges, implying lower variability and greater robustness. This improvement aligns with the results reported in [32], which established a hybrid reinforcement learning framework to enhance STAR-RIS beamforming. The authors of [32] claim that the convergence rate and stability of hybrid learning processes can be significantly enhanced by addressing the shortcomings of individual DRL algorithms .Unlike [31], which focused on improving stability within a single-agent TD3 framework, and [32], which combined hybrid learning into STAR-RIS systems, the proposed framework advances the field by incorporating hybrid decision-making processes tailored to RIS-assisted 6G IoT environments. Figure 2 demonstrates that the Best-Action and Dynamic-α(t) hybrid policies produce enhanced episode-based returns and display increased stability in learning behavior, particularly in environments prone to interference and marked by fast propagation shifts .The comparison with [31] and [32] illustrates that, unlike previous studies that emphasize specific deficiencies of individual DRL methods, our proposed hybrid framework concurrently improves return maximization, training stability, and resilience. The aforementioned qualities make the proposed method particularly suitable for practical implementation in the upcoming 6G IoT networks.

Figure 5 juxtaposes the learning behaviors of the assessed policies across the training episodes. Although TD3 demonstrates greater stability than DDPG, both approaches exhibit performance variability in dynamic channel and interference settings. The suggested hybrid decision procedures demonstrate accelerated convergence and consistently superior returns in the final training period.

Table 2 presents a quantitative assessment of the convergent performance, utilizing the mean return and the 95% confidence interval for the final K episodes. The Hybrid Best-Action policy achieves the highest mean return with the tightest confidence interval, validating its exceptional stability and reliability. The alignment between the visual patterns in Figure 5 and the statistical outcomes in Table 2 corroborates the efficacy of the suggested hybrid framework.

These findings align with recent studies on DRL-based RIS. Yuan et al. [31] demonstrated that single-agent DRL can effectively optimize RIS in dynamic settings. In contrast, Jian et al. [32] highlighted the pronounced susceptibility of RIS-assisted systems to channel fluctuations and hardware characteristics. The proposed hybrid

framework promotes resilience by integrating complementary learning tendencies, resulting in improved convergence and more stable performance.

Figure 3 depicts the rolling standard deviation of episode returns to assess training stability. The findings demonstrate that DDPG exhibits significant variance, whereas TD3 offers some improvement but remains susceptible to dynamic disturbances, as noted in references [24] and [25]. The proposed hybrid policy exhibits consistently lower variation by integrating the exploration capabilities of DDPG with the precise value estimation of TD3, hence ensuring more stable and predictable learning behavior in RIS-assisted wireless contexts.

In short, the analysis of Figures 1 to 5 shows that the proposed hybrid DRL architecture improves coverage performance, accelerates convergence, enhances system stability and strikes the best balance between exploration and exploitation. The hybrid RIS optimization method works best in 6G IoT settings, which are known for their slow transitions and susceptibility to interference. Sometimes, DRL methods alone won't be enough to fix the problem

The proposed Hybrid Best-Action reinforcement learning framework for optimizing RIS-assisted beamforming performs effectively, as demonstrated by both experimental and statistical evidence. By systematically incorporating a variety of learning behaviors, the proposed methodology improves decision-making resilience in dynamic channel environments, unlike traditional single-policy deep reinforcement learning systems. The key enhancements over DDPG, along with the evidence that it performs as well as TD3, indicate that hybrid action selection can boost stability and performance without adding complexity to the model.

The ANOVA and power analysis demonstrate that the performance improvements we've observed are statistically significant and not attributable to sample variation. The proposed strategy's real-world usefulness is underscored by its large effect size compared with traditional baselines, especially when the goals are to stabilize coverage and optimize constant rewards.

Taking a systemic approach, the immediate advantages of stabilizing rewards are improved beam alignment reliability, more consistent signal coverage, and reduced performance variability in RIS-assisted 6G IoT environments. The proposed hybrid DRL approach is a significant advancement in improving the performance of intelligent and adaptive wireless networks.

We chose DDPG and TD3 because they are effective at solving optimisation problems with continuous actions in wireless environments aided by RIS. The presented work intentionally focuses on deterministic continuous-control DRL algorithms, as the RIS phase optimisation naturally constitutes a continuous decision-making problem. It is important to further enrich the comparative analysis by exploring other DRL variants, such as PPO and SAC, which represent an important direction for future research.

9. CONCLUSION AND FUTURE WORK

This study evaluated the effectiveness of hybrid deep reinforcement learning methods in enhancing beamforming in dynamic wireless environments facilitated by RIS. An extensive experimental evaluation was conducted to analyze conventional Deep Reinforcement Learning baselines (DDPG and TD3) in conjunction with several hybridization techniques, including fixed-weight, dynamic-weight, and best-action selection algorithms. The results show that the Hybrid Best-Action method yields the highest average return and is more stable than the other methods examined. Statistical validation shows that this method works much better than DDPG and has a big effect size, while still being competitive with TD3. The one-way ANOVA test shows that the differences in performance among the policies are not random; they are real. A post hoc power analysis indicates that the sample size is sufficient to detect moderate-to-large improvements in performance.

The Hybrid Best-Action method makes it easier to estimate how well something will work and reduces randomness. This makes beamforming more reliable and ensures that IoT networks using RIS achieve better coverage. Next-generation 6G systems require specialized optimization frameworks that can accommodate fluctuating channel conditions, user mobility, and energy-efficiency demands while maintaining statistical integrity.

This study demonstrates that structured hybrid reinforcement learning techniques, enhanced with statistical analysis, significantly outperform conventional single-policy deep reinforcement learning methods for optimizing smart wireless networks.

Future research directions include improving the proposed hybrid framework for scenarios with many users and multiple RIS, exploring ways to make large-antenna designs more scalable, and incorporating energy-efficiency constraints into the optimization objectives. Federated or distributed learning techniques may enhance the adaptability of decentralized 6G systems.

The proposed system, in contrast to traditional ensemble learning methods, features an adaptive decision-level hybridization mechanism that dynamically selects or amalgamates DRL policies based on reward stability and performance assessment. This organized action-level fusion diminishes reward variance while maintaining convergence speed, providing a balanced exploration-exploitation trade-off designed for interference-prone 6G IoT environments. The work thus transcends policy aggregation by delivering statistically proven stability improvements in continuous-control RIS optimization, Channel models, including ray-tracing and DeepMIMO datasets, in forthcoming implementations.

ACKNOWLEDGMENT

The author sincerely thanks Prof. Khalid Hamid Bilal for his dedicated supervision, technical advice, and valuable critiques, all of which have greatly enhanced the quality and rigour of this work.

REFERENCES

1. S. Shukla, Mohd. F. Hassan, D. C. Tran, R. Akbar, I. V. Paputungan, and M. K. Khan, 'Improving latency in Internet-of-Things and cloud computing for real-time data transmission: a systematic literature review (SLR)', Cluster Comput, vol. 26, no. 5, pp. 2657-2680, Oct. 2023, https://doi.org/10.1007/s10586-021-03279-3
2. 'The Evolution of Mobile Communication: A Comprehensive Survey on 5G Technology', J Sen Net Data Comm, vol. 4, no. 1, pp. 01-11, Mar. 2024, https://doi.org/10.33140/JSNDC.04.01.06
3. N. Lassoued and N. Boujnah, 'A Comprehensive Review of Energy Efficiency in 5G Networks: Past Strategies, Present Advances, and Future Research Directions', Computers, vol. 15, no. 1, Jan. 2026, https://doi.org/10.3390/computers15010050
4. K. Dulaj, A. Alhammadi, I. Shayea, A. A. El-Saleh, and M. Alnakhli, 'Harnessing Machine Learning for Intelligent Networking in 5G Technology and Beyond: Advancements, Applications and Challenges', IEEE Open Journal of Intelligent Transportation Systems, vol. 6, pp. 605-633, 2025, https://doi.org/10.1109/OJITS.2025.3564361
5. Z. Li, J. Wang, S. Zhao, Q. Wang, and Y. Wang, 'Evolving Towards Artificial-Intelligence-Driven Sixth-Generation Mobile Networks: An End-to-End Framework, Key Technologies, and Opportunities', Applied Sciences, vol. 15, no. 6, Mar. 2025, https://doi.org/10.3390/app15062920
6. V. Chamola, M. Shall Peelam, M. Guizani, and D. Niyato, 'Future of Connectivity: A Comprehensive Review of Innovations and Challenges in 7G Smart Networks', IEEE Open Journal of the Communications Society, vol. 6, pp. 3555-3613, 2025, https://doi.org/10.1109/OJCOMS.2025.3560035
7. S. Prasad Tera, R. Chinthaginjala, G. Pau, and T. Hoon Kim, 'Toward 6G: An Overview of the Next Generation of Intelligent Network Connectivity', IEEE Access, vol. 13, pp. 925-961, 2025, https://doi.org/10.1109/ACCESS.2024.3523327
8. S. Alraih et al., 'Revolution or Evolution? Technical Requirements and Considerations towards 6G Mobile Communications', Sensors, vol. 22, no. 3, Jan. 2022, https://doi.org/10.3390/s22030762
9. M. Chafii, L. Bariah, S. Muhaidat, and M. Debbah, 'Twelve Scientific Challenges for 6G: Rethinking the Foundations of Communications Theory', IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 868-904, 2023, https://doi.org/10.1109/COMST.2023.3243918
10. F. Zhu et al., 'Wireless Large AI Model: Shaping the AI-Native Future of 6G and Beyond', Dec. 18, 2025, arXiv: arXiv:2504.14653, https://doi.org/10.48550/arXiv.2504.14653
11. S. Shafaei et al., 'Toward AI in 6G: Concepts, Techniques, and Standards', IEEE Access, vol. 13, pp. 143843-143874, 2025, https://doi.org/10.1109/ACCESS.2025.3595752

12. M. R. Fasihi and B. L. Mark, 'Device-to-Device Communication in 5G/6G: Architectural Foundations and Convergence with Enabling Technologies', Jul. 09, 2025, arXiv: arXiv:2507.06946, https://doi.org/10.48550/arXiv.2507.06946.
13. Z. Aasa, F. Elias, and S. C. Ekpo, 'Hybrid Energy and Spectrum Efficient Wireless Network Design for 5G/6G And Wi-Fi 7/8 Applications', Jan. 08, 2026, Research Square. https://doi.org/10.21203/rs.3.rs-8535275/v1
14. H. Wang et al., 'Navigating the Dual-Use Nature and Security Implications of Reconfigurable Intelligent Surfaces in Next-Generation Wireless Systems', IEEE Communications Surveys & Tutorials, vol. 28, pp. 3346-3387, 2026, https://doi.org/10.1109/COMST.2025.3621610
15. W. M. Othman et al., 'Key Enabling Technologies for 6G: The Role of UAVs, Terahertz Communication, and Intelligent Reconfigurable Surfaces in Shaping the Future of Wireless Networks', Journal of Sensor and Actuator Networks, vol. 14, no. 2, Mar. 2025, https://doi.org/10.3390/jsan14020030
16. X. Gan et al., 'Multi-Functional Programmable Metasurfaces for 6G and Beyond', Dec. 07, 2025, arXiv: arXiv:2512.06693, https://doi.org/10.48550/arXiv.2512.06693.
17. A. Tishchenko et al., 'The Emergence of Multi-Functional and Hybrid Reconfigurable Intelligent Surfaces for Integrated Sensing and Communications - A Survey', IEEE Communications Surveys & Tutorials, vol. 27, no. 5, pp. 2895-2936, Oct. 2025, https://doi.org/10.1109/COMST.2024.3519785
18. M. I. Khalil, K. Wang, J. Lin, and J. Choi, 'Mitigating Phase Errors to Improve Signal Quality in RIS-Assisted Satellite Communications', IEEE Transactions on Vehicular Technology, vol. 74, no. 9, pp. 14388-14403, Sep. 2025, https://doi.org/10.1109/TVT.2025.3566480
19. M. Ejaz, G. Jinsong, M. Asim, K. A. Shakil, and M. A. Wani, 'Joint Phase-Shift and Power Allocation Optimization in RIS-Enhanced Wireless Networks: An Intelligent Framework', IEEE Open Journal of the Communications Society, vol. 6, pp. 7389-7404, 2025, https://doi.org/10.1109/OJCOMS.2025.3602856
20. M. Iqbal, T. Ashraf, M. Zubair, S. M. Jameel, M. Jazib, and J.-Y. Pan, 'A comprehensive survey on reconfigurable intelligent surfaces (RIS) and STAR-RIS for next-generation wireless networks', Discov Appl Sci, vol. 7, no. 11, p. 1253, Oct. 2025, https://doi.org/10.1007/s42452-025-07684-w
21. A. Taha, M. Alrabeiah, and A. Alkhateeb, 'Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning', IEEE Access, vol. 9, pp. 44304-44321, 2021, https://doi.org/10.1109/ACCESS.2021.3064073
22. M. Jian et al., 'Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques', Intelligent and Converged Networks, vol. 3, no. 1, pp. 1-32, Mar. 2022, https://doi.org/10.23919/ICN.2022.0005
23. B. Sheen, J. Yang, X. Feng, and M. M. U. Chowdhury, 'A Deep Learning Based Modeling of Reconfigurable Intelligent Surface Assisted Wireless Communications for Phase Shift Configuration', IEEE Open Journal of the Communications Society, vol. 2, pp. 262-272, 2021, https://doi.org/10.1109/OJCOMS.2021.3050119
24. S. Lu et al., 'Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Cooperative Jamming Model Design', IEEE Access, vol. 11, pp. 98764-98775, 2023, https://doi.org/10.1109/ACCESS.2023.3312546
25. Z. U. A. Tariq, E. Baccour, A. Erbad, and M. Hamdi, 'Reinforcement Learning for Resilient Aerial-IRS Assisted Wireless Communications Networks in the Presence of Multiple Jammers', IEEE Open Journal of the Communications Society, vol. 5, pp. 15-37, 2024, https://doi.org/10.1109/OJCOMS.2023.3334489
26. T. Ma, Y. Xiao, X. Lei, L. Zhang, Y. Niu, and G. K. Karagiannidis, 'Reconfigurable Intelligent Surface-Assisted Localization: Technologies, Challenges, and the Road Ahead', IEEE Open Journal of the Communications Society, vol. 4, pp. 1430-1451, 2023, https://doi.org/10.1109/OJCOMS.2023.3292052

27. J. Wang and S. Chen, 'Deep Reinforcement Learning-Based Secrecy Rate Optimization for Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface-Assisted Unmanned Aerial Vehicle-Integrated Sensing and Communication Systems', Sensors, vol. 25, no. 5, Mar. 2025, https://doi.org/10.3390/s25051541
28. J. Chen et al., 'Hybrid Reinforcement Learning for Joint Beamforming in STAR-RIS-Assisted CoMP Systems', IEEE Transactions on Wireless Communications, vol. 24, no. 9, pp. 7955-7969, Sep. 2025, https://doi.org/10.1109/TWC.2025.3563597
29. Y. Zhang et al., 'A Unified Deterministic Channel Model for Multi-Type RIS With Reflective, Transmissive, and Polarization Operations', IEEE Transactions on Vehicular Technology, pp. 1-13, 2025, https://doi.org/10.1109/TVT.2025.3605727
30. Y. Huang et al., 'Sum Rate Maximization in STAR-RIS-UAV-Assisted Networks: A CA-DDPG Approach for Joint Optimization', Dec. 01, 2025, arXiv: arXiv:2512.01202. https://doi.org/10.48550/arXiv.2512.01202.
31. C. Cai, X. Yuan, W. Yan, Z. Huang, Y.-C. Liang, and W. Zhang, 'Hierarchical Passive Beamforming for Reconfigurable Intelligent Surface Aided Communications', IEEE Wireless Communications Letters, vol. 10, no. 9, pp. 1909-1913, Sep. 2021, https://doi.org/10.1109/LWC.2021.3085497
32. R. Zhong, Y. Liu, X. Mu, Y. Chen, X. Wang, and L. Hanzo, 'Hybrid Reinforcement Learning for STAR-RISs: A Coupled Phase-Shift Model Based Beamformer', IEEE Journal on Selected Areas in Communications, vol. 40, no. 9, pp. 2556-2569, Sep. 2022, https://doi.org/10.1109/JSAC.2022.3192053
33. X. Yuan, S. Hu, W. Ni, X. Wang, and A. Jamalipour, 'Deep Reinforcement Learning-Driven Reconfigurable Intelligent Surface-Assisted Radio Surveillance with a Fixed-Wing UAV', IEEE Transactions on Information Forensics and Security, vol. 18, pp. 4546-4560, 2023, https://doi.org/10.1109/TIFS.2023.3297021
34. https://www.kaggle.com/datasets/ziya07/6g-iot-intelligent-management-dataset
35. 'A machine learning approach to assess the climate change impacts on single and dual-axis tracking photovoltaic systems | Scientific Reports'. Accessed: Feb. 10, 2026. [Online]. Available: https://www.nature.com/articles/s41598-025-10831-3

ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING *Abdul Khalid and Alam*
Vol. 6 No. 1 2026
E-ISSN: 2785-8189

Analysis of Variable Frequency Drive for Electric Vehicles

Mohamad Izzat Za'im bin Abdul Khalid* and A. H. M. Zahirul Alam

[1]*Dept. of Electrical and Electronics Engineering, Faculty of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*

*Corresponding author: izzatzaim@gmail.com

(Received: 15 May 2026; Accepted: 16 June 2026)

Abstract— Efficient power electronics are critical in electric vehicle powertrains to optimize motor controls. Traditional traction topologies suffer from high switching losses and design complexities due to intermediate DC-DC boost stages. This paper evaluates the operational boundaries of a three-phase Variable Frequency Drive. Developed within the PSpice, the simulation maps the structural trade-offs between 180° and 120° conduction topologies under variable resistance-inductance load lines. Results show that while the 180° mode maintains line voltage stability across all loads, it poses a severe risk of phase-leg shoot-through short circuits during switching intervals. Conversely, the 120° mode prevents shoot-through by providing an inherent 60° non-conducting safety window. The parametric sweeps show that under a heavy inductive load line, the 120° voltage waveform collapses into an asymmetric triangular profile. Furthermore, transient testing indicates that open-loop Piecewise Linear modulations result in severe gate pulse overlap. These findings establish critical boundary constraints vital for deploying secure vehicle inverter drive control loops.

Keywords: *Conduction topologies, electric vehicle, variable frequency drive*

1. INTRODUCTION

The global transition toward electric vehicles (EVs) requires high-efficiency power electronic traction drives to maximize vehicle range and system efficiency [1], [2]. Conventional architectures employ a dual-stage layout, using a standalone intermediate DC-DC boost converter to step up the battery terminal voltage before feeding a three-phase Variable Frequency Drive (VFD). While effective for voltage scaling, this separate boost stage increases design complexities and high-frequency switching losses. To mitigate these hardware integration constraints, single-stage three-phase VFDs are a reliable alternative for directly interfacing the DC source to AC traction motors [3] [4]. In these single-stage configurations, the choice of semiconductor gating topology dictates the hardware's structural reliability and the quality of the output power [5]. The two foundational control sequences used to trigger the bridge semiconductor switches are the 180° and 120° conduction modes [7]. Both configurations exhibit distinct operational characteristics: the 180° mode maximizes DC rail utilization, whereas the 120° mode introduces a non-conducting interval to prevent simultaneous phase-leg conduction [7].

A critical gap in existing power electronics literature is the lack of detailed comparative mapping regarding how these fundamental conduction modes behave under varied load impedances under open-loop conditions. Most contemporary research focuses immediately on complex closed-loop algorithms such as Field-Oriented Control (FOC) or Space Vector Pulse Width Modulation (SVPWM), bypassing the baseline hardware constraints and transient behaviors. This paper addresses the identified research gap by establishing an operational boundary framework for both the 180° and 120° conduction topologies. Using PSpice simulations, this study evaluates the structural behavior of a three-phase inverter bridge subjected to multi-tiered resistance-inductance (R-L) load lines, variable operating-frequency channels, and altered gating pulse widths.

2. BACKGROUND OF THE STUDY

To evaluate the structural trade-offs of a single-stage three-phase VSI, the underlying physics of the semiconductor triggering sequences must be established. The inverter bridge utilizes six voltage-controlled

DOI: https://doi.org/10.69955/ajoeee.2026.v6i2.87

switches operating in a complementary sequence across three-phase legs (A, B, C) as shown in Fig. 1 . The output line-to-line voltage equations and current pathways are directly dictated by the conduction interval allocated to each active switch [6].

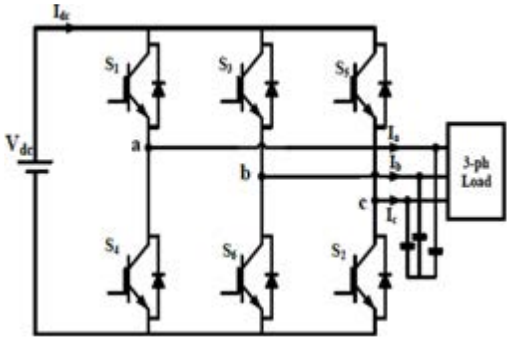

Fig. 1. Circuit diagram of 3-phase VFD

2.1 180° Conduction Mode

In the 180° conduction configuration, each semiconductor switch conducts for a full half-cycle duration. At any given instance, exactly three switches across the bridge remain active simultaneously. The switching sequence changes at 60° intervals, generating a balanced three-step quasi-square line-voltage waveform. The primary mathematical advantage of this mode is its high efficiency in utilizing the DC rail Vdc input. However, when upper and lower switches are on the same phase leg transition instantaneously, the topology introduces a critical vulnerability to phase-leg shoot-through short circuits if gating pulse timings overlap during dead-time failures [6]. The timing diagram is shown in Fig. 2.

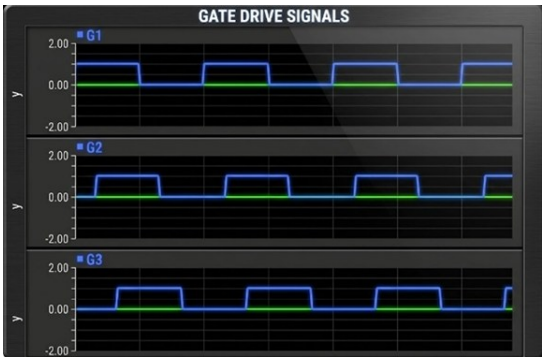

Fig. 2. 180° mode pulse

2.2 120° Conduction Mode

In the 120° conduction configuration, each active switch is gated to conduct for exactly one-third of the total cycle period, as shown in Fig. 3. Therefore, only two switches conduct simultaneously at any single operating window [8]. This mode enforces an inherent 60° non-conducting window per cycle, providing a structural 30° safety dead-time margin. While this safety dead time inherently eliminates the risk of phase-leg shoot-through, the output line voltage assumes a six-step stepped waveform that is highly dependent on the connected load line [3].

2.3 Inductive Load Interactions and Wave Deformation

When feeding a balanced delta-connected resistance-inductance (R-L) load, the current waveform does not instantly follow the voltage steps due to inductive reactance. The stored magnetic energy must be discharged through the anti-parallel freewheeling diodes during unexcited intervals [8]. Under extreme inductive conditions, this continuous freewheeling current discharge forces a geometric collapse of the active voltage steps, shifting the envelope from a staircase profile to an asymmetric triangular wave, thereby inducing high-order harmonic distortions in traction systems [4].

Fig. 3. 120° mode pulse

3. METHODOLOGY

The paper identifies variable-frequency drive (VFD) requirements and then develops the three-phase inverter bridge circuit layout using Capture software. This schematic configuration is subsequently imported into PSpice to execute time-domain transient simulations. Within the simulation environment, multi-variable parametric sweeps are systematically conducted to evaluate different gating conduction modes and alternative modulation coordination methods across variable resistance-inductance (R-L) load-line configurations. Finally, the simulated line-to-line output voltage waveforms are extracted and evaluated to assess signal integrity, underdamped switching transients, and geometric variations before considering the overall operational boundaries and hardware constraints.

The simulation framework was designed in the PSpice environment to analyze the transient and steady-state performance of a three-phase voltage-source inverter (VSI) bridge driving a balanced R-L load, as shown in Fig. 4. The schematic architecture consists of six voltage-controlled semiconductor switches modeled with low on-state resistance parameters to simulate a Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) [6]. Each active switch is paired with an antiparallel freewheeling diode to permit inductive current to circulate and ensure safe energy dissipation during transition intervals. The DC input rail (Vdc) is maintained at a constant 250V baseline to energize the load network.

3.1 Gate Pulse Sequences

Independent pulse generators control the MOSFET triggering profiles, synchronized to a precise 120° phase shift across the three-phase output legs (A, B, C). Both modes are tested with a variation of loads.

- *180° Conduction Mode:* Controlled via pulse widths set to exactly 10 ms at a 50 Hz frequency, ensuring continuous gating over half the operational period.
- *120° Conduction Mode:* Regulated using restricted pulse widths of 6.67 ms within the same period, establishing the structural 3.33 ms non-conducting gap between upper and lower devices.

3.2 Parametric Sweep Configuration

To evaluate the stability boundaries of the gating topologies under dynamic operating conditions, multi-variable parametric sweeps are performed in PSpice. The first sweep channel deviates the fundamental operating frequency across three discrete steps: 40 Hz, 50 Hz (nominal baseline), and 55 Hz. The second independent sweep channel scales the active gating pulse width from 6 ms to 12 ms against the 10 ms nominal timing. These sweeps are automated to systematically assess the tolerance of output voltage envelopes while holding other parameters constant.

3.1 Piecewise Linear (VPWL) Setup

An independent modulation channel is established by replacing the periodic pulse generators with open-loop Piecewise Linear (VPWL) voltage sources. This configuration uses precise time-voltage coordinate pairs in the

PSpice property matrix to define custom, non-uniform switching transitions. The VPWL setup is implemented to test the synchronization limits of the inverter bridge and to investigate how the open-loop system reacts to non-periodic gating signals.

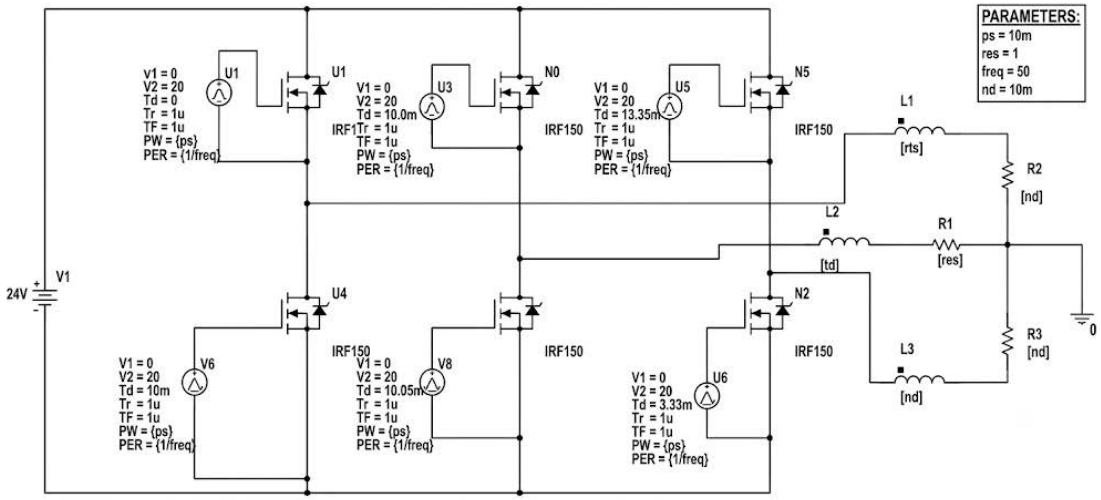

Fig. 4. Schematic diagram of the three-phase VSI connected to R-L load

4. RESULT AND ANALYSIS

4.1 Evaluation of Conduction Modes Under Variable Load Impedance

The time-domain simulation waveforms indicate geometric stability between the two conduction topologies as the load line reactance varied. In the 180° conduction mode, the line-to-line output voltage (Vab) consistently tracks a stable, balanced three-step quasi-square-wave profile as shown in Fig. 5. This geometric envelope is preserved across all testing quadrants from Case 1 through Case 4, proving that the continuous 180° firing pattern is highly immune to variations in load impedance. However, because complementary upper and lower switches on the same phase leg transition instantaneously without an implicit delay, this topology poses a severe risk of phase-leg shoot-through short circuits if gating pulses overlap, matching the safety vulnerabilities [6].

Fig. 5. Line-to-line output voltage (Vab) waveform under 180° conduction mode for Case 1.

The 120° conduction mode exhibits extreme sensitivity to load variations. In Case 1, the line voltage yields a clean staircase pattern with minor geometric distortion, providing a smoother transition that approaches a sinusoidal envelope as depicted in Fig. 6. When subjected to Case 2 (R = 10 Ω, L = 1 mH), severe switching instability occurs, where the line voltage exhibits high-amplitude peak voltage spikes at every switching edge as shown in Fig. 6. Under Case 3 (R = 1 Ω, L = 10 mH), the increased inductive reactance acts as a natural filter,

damping these transient spikes to yield an ideal staircase-step envelope. However, under Case 4 (R = 10 Ω, L = 10 mH), the active voltage steps become completely deformed and collapse into an asymmetric triangular waveform as observed in Fig. 7. This structural degradation occurs because the heavy inductive load stores significant reactive energy, forcing continuous current to flow through the antiparallel freewheeling diodes during the designated 60° non-conducting window. nder heavy inductive loads, the active voltage steps collapse into an asymmetric triangular waveform, confirming structural degradation [3].

Fig. 6: Three-phase line voltage waveforms in 120° conduction mode under (Case 1)

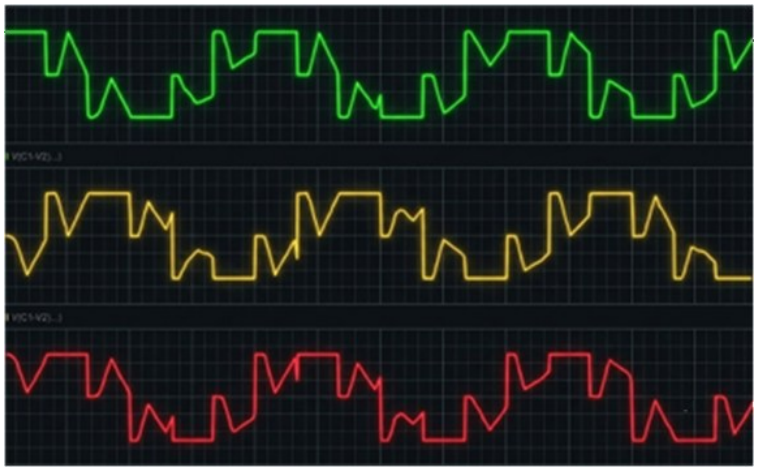

Fig. 7 Geometric collapse of line-to-line voltage (Vab) into a triangular profile under 120° conduction mode for Case 4.

4.2 Parametric Sweeps of Frequency and Pulse Width Channels

Deviating the operating frequency from the nominal 50 Hz baseline significantly limits power quality across variable inductive loads (L = 1 mH to L = 100 mH). Under the low-inductance profile (L = 1 mH), operating at the lower channels of 40 Hz and 45 Hz induces minor transient ringing along the switching edges as observed in Fig. 8. When the inductance is stepped up to 10 mH, an overall improvement is observed, with the waveform becoming exceptionally smooth at 50 Hz and 55 Hz. However, under a heavy load inductance of 100 mH, the output waveform at the 40 Hz channel exhibits severe transient ringing along the transition boundaries, indicating an unstable operating zone that the drive system must actively avoid.

Varying the gating pulse width shows that shorter durations (6 ms and 8 ms) reduce the effective root-mean-square (RMS) output voltage, producing clean but narrow voltage blocks that lower torque production. At the nominal 10 ms duration, the waveform yields a highly stable, symmetrical three-step quasi-square profile under baseline conditions. However, expanding the pulse width to 12 ms under low inductance forces the waveform to experience severe transient ringing at the switching boundaries, making it an unsatisfactory configuration. In contrast, under higher load conditions (L = 10 mH and L = 100 mH), the 120 ms expanded width successfully

widens the active steps while perfectly preserving a uniform, symmetrical transition with minimal distortion.

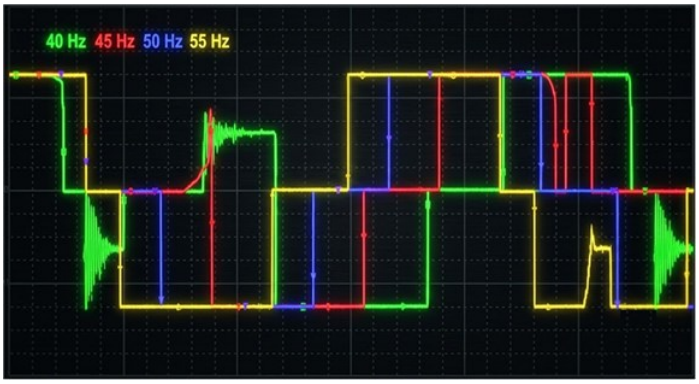

Fig. 8: Line voltage characteristics under frequency scaling (40 Hz - 55 Hz) with 100 mH

Varying the gating pulse width shows that shorter durations (6 ms and 8 ms) reduce the effective root-mean-square (RMS) output voltage, producing clean but narrow voltage blocks that lower torque production. At the nominal 10 ms duration, the waveform yields a highly stable, symmetrical three-step quasi-square profile under baseline conditions. However, expanding the pulse width to 12 ms under low inductance forces the waveform to experience severe transient ringing at the switching boundaries, making it an unsatisfactory configuration. In contrast, under higher load conditions (L = 10 mH and L = 100 mH), the 120 ms expanded width successfully widens the active steps while perfectly preserving a uniform, symmetrical transition with minimal distortion as observed in figures 9 and 10.

Fig. 9 Line voltage characteristics under pulse width (6 ms – 12 ms) with 1 mH

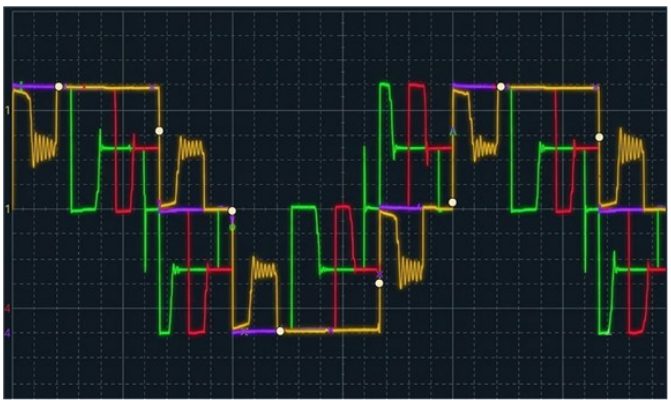

Fig. 10: Line voltage characteristics under pulse width (6 ms – 12 ms) with 100 mH

4.3 Evaluation of VPWL Modulation Stability

The implementation of coordinate-based VPWL inputs without a closed-loop feedback network proved highly ineffective for three-phase drive control. Due to the lack of strict time synchronization during non-uniform transitions, the independent gating pulses experienced severe phase overlap and timing collisions. This structural breakdown triggered continuous, high-frequency edge noise and heavy higher-order harmonic injection, confirming that open-loop VPWL controls are entirely unsuitable for secure vehicle traction operations, as observed in Fig. 12.

Fig. 12: Gate control signal overlaps and waveform switching collisions under VPWL coordination

5. CONCLUSION

This study has successfully mapped the strict operational boundaries, performance trade-offs, and geometric deformation limits of three-phase VSI gating strategies for EV applications using PSpice. The 180° mode provides superior line-voltage stability under dynamic loads but poses severe shoot-through risks. The 120° mode provides an inherent safety dead-time window that eliminates shoot-through paths, and VPWL coordination is verified as unviable due to gating overlaps. Future research will focus on integrating closed-loop Proportional-Integral-Derivative (PID) feedback networks to effectively minimize total harmonic distortion (THD) under high-power EV drive demands.

REFERENCES

[1] A. Athwer and A. Darwish, "A review on modular converter topologies based on WBG semiconductor devices in wind energy conversion systems," Energies, vol. 16, no. 14, art. no. 5324, Jul. 2023, doi: 10.3390/en16145324.

[2] M. H. Nguyen and S. Kwak, "Enhance reliability of semiconductor devices in power converters," Electronics, vol. 9, no. 12, art. no. 2068, Dec. 2020, doi: 10.3390/electronics9122068.

[3] M. M. Qasim, D. M. Otten, J. H. Lang, J. L. Kirtley, and D. J. Perreault, "Comparison of inverter topologies for high-speed motor drive applications," IEEE Trans. Power Electron., vol. 39, pp. 7404–7422, 2024, doi: 10.1109/tpel.2024.3376196.

[4] S. K. K. Sampathkumar and D. Pradyumna Kumar, "Power quality assessment of novel multilevel and multistring inverters for electric vehicle applications," Bull. Electr. Eng. Inform., vol. 11, no. 4, pp. 1818–1827, Aug. 2022, doi: 10.11591/eei.v11i4.3512.

[5] T. Saravanakumar and R. Saravana kumar, "Design, validation, and economic behavior of a three-phase interleaved step-up DC–DC converter for electric vehicle application," Front. Energy Res., vol. 10, art. no. 813081, 2022, doi: 10.3389/fenrg.2022.813081.

[6] R. Shweta, S. Sivagnanam, and K. A. Kumar, "Fault detection and monitoring of solar photovoltaic panels using internet of things technology with fuzzy logic controller," Electr. Eng. Electromech., no. 6, pp. 67–74, 2022, doi: 10.20998/2074-272x.2022.6.10.

[7] G. Susinni, S. A. Rizzo, and F. Iannuzzo, "Two decades of condition monitoring methods for power devices," Electronics, vol. 10, no. 6, art. no. 683, Mar. 2021, doi: 10.3390/electronics10060683.

ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING
Vol. 6 No. 1 2026
E-ISSN: 2785-8189

TD3 vs. DDPG for RIS-Assisted Beamforming Optimization: Statistical and Communication-Level Analysis for 6G IoT Networks

Ezdihar Osman Taj Almowla Mohomad[1*], Khalid Hamid Bilal[2], Zeinab Mahmoud Omer[1], Abeer Mohamed Elzain[3], and Rania Ali Elkhidir[4]

[1]*University of Bahri Khartoum, Sudan*
[2]*University of Science & Technology, Omdurman*, Sudan
[3]*University of Bahri, Khartoum. Sudan*
[4]*University of Hail, Saudi Arabia*

*Corresponding author: Ezdiharosman22@gmail.com

(Received: 23 February 2026; Accepted: 14 June 2026)

Abstract—Reconfigurable Intelligent Surfaces (RIS) are expected to play a critical role in future 6G wireless networks by enabling adaptive and intelligent signal propagation. This study presents a deep reinforcement learning (DRL)-based simulation framework for beamforming optimization in RIS-assisted wireless communication systems. Two continuous-control DRL algorithms, namely Twin Delayed Deep Deterministic Policy Gradient (TD3) and Deep Deterministic Policy Gradient (DDPG), were evaluated under 30 independent wireless channel scenarios. The experimental results demonstrate that TD3 achieves a higher average episodic return (13.34 ± 0.73) than DDPG (11.13 ± 0.87), representing an improvement of approximately 19.86%. Statistical analysis further confirms that this improvement is significant ($t \approx 3.99$, $p < 0.001$) with a large effect size (Cohen's $d = 1.03$). The 95% confidence interval of the mean difference ranges from 1.10 to 3.32. In addition to better average performance, TD3 exhibits steadier convergence and reduced reward variability. This means that learning is more robust under dynamic wireless channel conditions. The results suggest that TD3 provides a robust and statistically reliable method for optimizing intelligent beamforming in RIS-assisted 6G IoT communication scenarios, thereby improving wireless connectivity performance and system stability.

Keywords: Reconfigurable Intelligent Surfaces (RIS), Coverage Enhancement,6G Internet of Things (6G IoT), Deep Deterministic Policy Gradient (DDPG), Deep Reinforcement Learning (DRL)

1. INTRODUCTION

A low-latency, comprehensive, fast, and reliable connection is now more important than ever, thanks to advances in wireless communication technology [1][2][3]. To meet these critical needs, the International Telecommunication Union (ITU) and other interested parties have developed a plan and primary goals for sixth-generation (6G) networks. Most people think that the most important thing is to strengthen the network, integrate the complete Internet of Things (IoT), and improve services [4] [5].

Included in the many traditional means of human-to-human contact are base stations [6][7][8]. Base stations have difficulty handling dense, broad IoT deployments due to transmission issues such as fluctuating interference [9][10], multipath fading, substantial obstruction, and instances of undetected signals (NLoS) [11][12]. To enhance wireless communication while reducing prices and energy usage, researchers have investigated novel methods. Researchers found that intelligent reflecting surfaces (IRS) and reconfigurable intelligent surfaces (RIS) can enhance wireless networks [13]. Researchers can construct a spectrum of RIS-based passive reflective devices to respond to various phases [14]. This implies the ability to instantly alter electromagnetic waves [15]. Reducing noise while increasing signal length and intensity is the goal of this function. Because it requires little power, no

DOI: https://doi.org/10.69955/ajoeee.2026.v6i1.84

costly equipment [16], and no complex installations, RIS is a crucial technology for 6G IoT communications [17].

With a small number of Internet of Things (IoT) devices and multiple low-power devices operating simultaneously [17] [18], traditional beamforming methods may rely on coverage or fail to meet quality-of-service (QoS) standards [19]. Not only is space navigation severely lacking, but the tube itself is woefully inadequate. By increasing the number of signal segments, decreasing interference, and creating links that appear to be in line of sight, RIS can overcome these issues. When it comes to improving and expanding the network, the invention is the way to go. We are still in the early stages of developing RIS beamforming for 6G IoT networks. Despite RIS's many obvious advantages [20], determining the optimal beamforming and phase-shift configurations for a large-scale Internet of Things system [21] can be difficult. The RIS provides a multilevel setting for continuous activity by allowing each reflecting element to select its phase from the range [0, 2π] [22]. The base station broadcasts a non-linear RIS-IoT channel that is highly linked. It changes over time, especially in busy or noisy environments [23]. Methods require substantial processing power, making immediate enhancement difficult. People rarely use heuristics such as semidefinite relaxation, alternating optimization, and search strategies [24]. These processes can be time-consuming, expensive, and inefficient. Furthermore, in real-world IoT applications, it can be difficult to generate highly accurate channel models or precise channel state information (CSI) that these methods often require [25]. When dealing with several variables, changing conditions, and ongoing supervision, intelligent optimization frameworks become indispensable. Deep Reinforcement Learning (DRL) offers a significant advantage for making complex decisions in unpredictable, rapidly changing environments [26]. Instead of relying on explicit channel models or trial and error, DRL automatically learns the best rules. Rather, it engages with its surroundings instantly. Deep Deterministic Policy Gradient (DDPG) and Twin Delayed Deep Deterministic Policy Gradient (TD3) are two actor-critic approaches that could improve systems requiring ongoing refinement. Beamforming and RIS can now work together.

Recent research has focused on specific areas such as security, anti-jamming, and UAV communication, as well as optimal network topologies and discrete or quantized action spaces. However, several RIS-based wireless systems have incorporated Deep Reinforcement Learning (DRL) [27]. Research on modern continuous-control deep reinforcement learning methods used in common RIS-IoT applications is lacking. Despite the importance of these challenges to the successful implementation of 6G, problems with training stability, convergence characteristics, and resilience to inaccurate or misleading Channel State Information (CSI) persist.

A continuous-control deep reinforcement learning strategy may enhance RIS-assisted beamforming in 6G IoT networks, according to the study's findings. The Markov Decision Process (MDP) allows beamforming and phase shifting in the RIS. Our empirical evaluations focused on DDPG and TD3, two of the most prominent actor-critic systems. To improve the signal-to-interference-plus-noise ratio (SINR) and network coverage, the suggested approach uses realistic 3D topology modeling and various IoT scenarios. Its size will have no effect on its effectiveness.

The main contributions of this work can be summarized as follows:

1. A single, continuous-action deep reinforcement learning system for optimizing beamforming in large 6G IoT networks with reconfigurable intelligent surface integration.
2. A complete MDP formulation that covers the state space, action space, reward function, and system dynamics of RIS-IoT communications .
   A detailed comparison of DDPG and TD3 in the same RIS-IoT setups, looking at how fast they converge, how stable they are, and how strong they are.
3. A method for making datasets that show 3D topology, describe the properties of wireless channels, set up BS-RIS-IoT, and spread users out.
4. A comprehensive performance evaluation that validates the effectiveness of the proposed strategy through convergence analysis, SINR distributions, and coverage maps.

Key Contributions and Novelty of the Proposed Framework

This work is novel in that it provides a comprehensive comparative study of the continuous-control DRL algorithms for RIS-assisted beamforming optimizations under the same wireless communication settings. Compared with traditional studies that mainly focus on reward convergence, the proposed approach provides a multi-dimensional evaluation, including reward stability, statistical distribution analysis, SINR performance, sum-

rate analysis, and convergence robustness.

More specifically, the present study contributes to the following:

- A unified RIS-assisted DRL evaluation framework for fair comparison between DDPG and TD3 under identical wireless channel conditions.
- A detailed statistical performance analysis including confidence intervals, variance analysis, boxplot evaluation, and effect-size interpretation.
- Communication-level performance evaluation using SINR and achievable sum-rate metrics in addition to reward-based learning analysis.
- A practical discussion on real-world RIS deployment challenges and future integration with realistic wireless propagation environments such as DeepMIMO and ray-tracing models.

The results obtained demonstrate that TD3 provides significantly improved convergence stability, reward consistency, and communication reliability compared to DDPG in dynamic RIS-assisted 6G IoT environments.

Recently, reinforcement learning algorithms, such as Soft Actor–Critic (SAC) and Proximal Policy Optimization (PPO), have shown strong performance in continuous control tasks. In this work, we focus on deterministic policy-gradient methods. We select DDPG and TD3 because TD3 was initially proposed as a direct improvement over DDPG, enabling a controlled, systematic comparison under the same RIS-assisted wireless communication conditions. We restrict the analysis to closely related deterministic actor-critic algorithms to isolate the effect of TD3's architectural enhancements: twin critics, delayed policy updates, and target-policy smoothing. This avoids introducing additional algorithmic differences related to entropy-based or policy-gradient approaches.

2. RELATED WORK

Recent research has increasingly explored deep reinforcement learning (DRL) techniques for reconfigurable intelligent surface (RIS)–assisted wireless systems. In [28], a DRL-based joint beamforming design was investigated for RIS-assisted wireless networks; however, the study relies on a single learning paradigm and focuses mainly on rate optimization rather than coverage enhancement. The work in [29] extended DRL to joint deployment and passive beamforming optimization in RIS-assisted networks, yet it did not examine algorithmic comparisons under a unified environment. A broader DRL-oriented perspective on RIS-assisted communications was presented in [30], which outlined key challenges and opportunities but lacked a concrete MDP formulation and comparative performance analysis. More recent studies, such as [31], addressed DRL-based beamforming optimization in RIS-assisted multi-user MISO systems, primarily targeting spectral efficiency and robustness under specific channel conditions. Additionally, TD3-based joint beamforming frameworks have been proposed for RIS-assisted systems with imperfect CSI in [32]. Still, these works employ a single DRL algorithm and do not focus on IoT-centric coverage metrics. Unlike the aforementioned studies, the present work formulates a complete MDP. It provides a unified comparative evaluation of DDPG and TD3 for coverage enhancement in dense RIS-assisted 6G IoT networks, thereby constituting a clear methodological and performance-driven advancement over existing literature.

Existing research on RIS-assisted DRL is largely limited to single-metric optimization or reward-convergence behaviors in constrained simulation environments. Furthermore, few previous studies have provided extensive statistical analysis, communication-level performance analysis, and comparative stability assessment of continuous-control DRL algorithms under the same wireless channel conditions. Moreover, the throughput-oriented evaluation and SINR-based coverage analysis have received less consideration in the context of RIS-assisted IoT communication. To address these limitations, the proposed work offers a comprehensive comparative approach to assess DDPG and TD3 by considering statistical analysis, reward-distribution interpretation, SINR evaluation, and achievable sum-rate performance under the same RIS-assisted wireless settings.

3. SYSTEM OVERVIEW AND PROBLEM FORMULATION

3.1 Network Model

As depicted in Fig. 1, we examine a three-dimensional (3D) RIS-assisted multi-user IoT downlink system functioning within a 6G framework. A base station (BS) with N_t antennas serves K single-antenna IoT users via a reconfigurable intelligent surface (RIS) comprising M passive reflecting elements.

Two links establish communication between the BS and IoT users: a direct link from the BS to the user and an indirect, reflected link via the RIS. The RIS facilitates coherent signal combining at the receivers by dynamically adjusting its phase shifts, which improves coverage and signal quality in dense IoT environments.

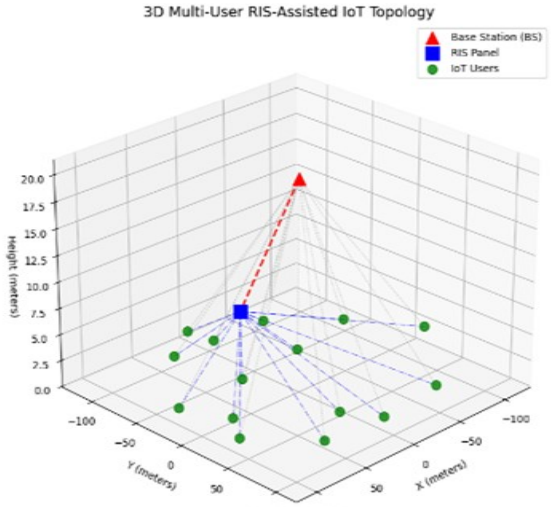

Fig.1 illustrates the 3D multi-user RIS-assisted IoT topology used in this study

3.2 Channel Model

The direct channel between the BS and the $\boldsymbol{k}$-th user is denoted by

$$\mathrm{h}_{b,k} \in \mathbb{C}^{1\times N_t}. \qquad (1)$$

The channel from the BS to the RIS and from the RIS to the $\boldsymbol{k}$-th user are denoted by

$$\mathrm{G}_{b,r} \in \mathbb{C}^{M\times N_t}, \mathrm{h}_{r,k} \in \mathbb{C}^{1\times M}, \quad (2)$$

respectively.

The RIS reflection matrix is defined as

$$\Theta = diag\left(e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_M}\right), \qquad (3)$$

where $\boldsymbol{\theta_m} \in [\mathbf{0}, \mathbf{2\pi})$represents the phase shift of the $\boldsymbol{m}$-th RIS element.

Accordingly, the effective channel for the $\boldsymbol{k}$-th user is given by

$$\mathrm{h}_k^{\mathrm{eff}} = \mathrm{h}_{b,k} + \mathrm{h}_{r,k}\Theta\mathrm{G}_{b,r}. \quad (4)$$

3.3 Signal Model

The transmitted signal from the base station (BS) is given by

$$\mathrm{x} = \sum_{k=1}^{K} \mathrm{w}_k\, s_k, \qquad (5)$$

where $w_k \in \mathbb{C}^{N_t \times 1}$denotes the beamforming vector associated with the k-th IoT user, and $s_k$represents the corresponding information symbol satisfying $\mathbb{E}[|\ s_k\ |^2] = 1$.

The received signal at the k- The user can be expressed as

$$y_k = h_k^{eff} w_k s_k + \sum_{i \neq k} h_k^{eff} w_i s_i + n_k, \qquad (6)$$

where h_k^{eff}denotes the effective BS–RIS–user channel, and $n_k \sim \mathcal{CN}(0, \sigma^2)$represents additive white Gaussian noise (AWGN) with variance σ^2.

Accordingly, the signal-to-interference-plus-noise ratio (SINR) at the k- The user is defined as

$$SINR_k = \frac{|\ h_k^{eff} w_k\ |^2}{\sum_{i \neq k} |\ h_k^{eff} w_i\ |^2 + \sigma^2}. \qquad (7)$$

3.4 Problem Formulation

The objective of this work is to jointly optimize the BS beamforming vectors. $\{w_k\}$and the RIS phase shift matrix ΘTo enhance coverage and communication quality in dense IoT deployments. The optimization problem is formulated as

$$\begin{aligned} \max_{\{w_k\}, \Theta} \quad & \sum_{k=1}^{K} \log_2 (1+SINR_k) \\ \text{s.t.} \quad & \sum_{k=1}^{K} \|\ w_k\ \|^2 \leq P_{max}, \\ & \theta_m \in [0, 2\pi), \forall m, \end{aligned} \qquad (8)$$

where P_{max}denotes the maximum transmit power at the BS. This optimization problem is highly non-convex due to the coupling between beamforming vectors and continuous RIS phase shifts. Therefore, it is reformulated as a Markov decision process (MDP) and solved using continuous-control deep reinforcement learning, as described in the following section.

The first constraint ensures practical, energy-efficient operation by limiting the base station's total transmit power to the maximum allowable value Pmax. The second limitation restricts each RIS phase-shifting element to the physical range $[0,2\pi)$, which represents the achievable phase adjustment capability of the RIS hardware

The optimization problem is non-convex because the beamforming vectors and RIS phase-shift variables are jointly coupled in the SINR expression. Consequently, obtaining a globally optimal solution through conventional optimization techniques is computationally challenging, particularly in dynamic multi-user wireless environments. Therefore, the problem is reformulated as a Markov Decision Process (MDP) and solved using continuous-control deep reinforcement learning.

3.5 Motivation for DRL-Based Design

The problem of configuring RIS and joint beamforming involves high dimensionality, strong nonlinearity, and continuous control variables, which render traditional model-based optimization methods impractical in dynamic 6G IoT environments. By allowing the learning agent to engage directly with the wireless environment, deep reinforcement learning (DRL) offers an effective solution for learning optimal control policies without dependence on explicit channel models or convex reformulations.

Note that the channel model described in this section is used to build the theoretical framework for RIS-assisted communications and to specify the underlying communication relationships. In contrast, the DRL environment used in this study is based on dataset-derived communication-performance

features that are used for practical training, state representation, and performance evaluation.

4. DRL-BASED JOINT BEAMFORMING AND RIS OPTIMIZATION

4.1 DRL Framework Overview

A deep reinforcement learning framework is adopted to address the non-convex joint beamforming and RIS optimization problem in RIS-assisted multi-user IoT networks. The BS is modeled as the learning agent, while the RIS-assisted wireless system represents the environment. At each decision step, the agent observes the current network state and selects continuous control actions to improve coverage and signal quality.

4.2 Markov Decision Process Formulation

The RIS-assisted beamforming problem is formulated as an MDP defined by the tuple. $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$:

State: The state at time step $t$captures the essential channel and configuration information and is defined as

$$st = \{xt, Tt, Lt, Et, \gamma t\}, \qquad (9)$$

where $\boldsymbol{xt}$ denotes the normalized dataset-derived network features at time step t, $\boldsymbol{Tt}$ represents the normalized throughput, $\boldsymbol{Lt}$ represents the normalized latency, $\boldsymbol{Et}$ denotes the normalized energy-consumption value, and $\boldsymbol{\gamma t}$ denotes the SINR-related performance indicator when available.

The state representation is constructed from processed communication-performance features extracted from the dataset rather than explicit raw channel-state information (CSI). Consequently, the DRL agent observes a compact representation of the wireless communication environment that captures throughput, latency, energy efficiency, SINR-related indicators, and other relevant network-performance characteristics required for optimization.

Action: The action selected by the agent jointly controls the BS beamforming and RIS configuration and is defined as

$$\mathrm{a_t} = \{\mathrm{w_t}, \theta_\mathrm{t}\}, \theta_\mathrm{m} \in [0,2\pi), \mathrm{m}, \qquad (10)$$

where $\mathrm{w}_t$denotes the BS beamforming vector and $\theta_t$represents the continuous RIS phase-shift vector.

4.3 Action Space Limitations

We restrict the action space to the range [−1, 1] across all action dimensions, which enables stable learning and valid control decisions. When interacting with the environment, the actor network's output actions are clipped to fall within predefined limits before being applied.

The bounded continuous-action representation improves training stability and prevents the generation of very large control values that could negatively impact the optimization process. This constraint also ensures a consistent exploration strategy during training and evaluation while maintaining feasible actions throughout the learning process.

The reward function is defined to maximize communication performance by improving throughput while minimizing latency and energy consumption, and is expressed as

$$rt = wthrTt - wlatLt - wenEt \qquad (11)$$

where Tt, Lt, Et are the normalised throughput, latency and energy-consumption values at time step t, respectively. The contribution of each performance metric is controlled by the weights w_thr, w_lat, and w_en. In this study, w_thr=1.0, w_lat=0.3, and w_en=0.2.

This reward formulation encourages the agent to perform actions that increase throughput while decreasing latency and energy consumption. The reward value is clipped to [-1, 1] to ensure stable learning during training.

In real-world 6G IoT dynamics, continuous control optimization is possible because the environment transitions to a new state s_(t+1) due to changes in the channel and the chosen action.

This multi-objective reward design enables the agent to balance spectral efficiency, delay performance, and

energy efficiency while optimizing its policy.

Actor-critic deep reinforcement learning (DRL) algorithms are used to address continuous action spaces and sequential decision-making. In this study, DDPG and TD3 are selected because they are specifically designed for continuous control optimization problems and have shown strong performance in high-dimensional wireless communication environments.

DDPG learns deterministic policies using an actor–critic architecture, which improves training stability. But it can lead to value overestimation during training. TD3 mitigates this limitation by using twin critic networks, delayed policy updates and target-policy smoothing. Such improvements alleviate overestimation bias and improve the robustness of the policy, making TD3 especially suitable for continuous-control optimisation problems.

4.4 DDPG and TD3-Based Learning

Actor–critic DRL techniques manage the continuous action space associated with RIS phase shifts and beamforming vectors. Specifically, the Deep Deterministic Policy Gradient (DDPG) and Twin Delayed Deep Deterministic Policy Gradient (TD3) are investigated.

DDPG directly learns deterministic policies for continuous control, whereas TD3 mitigates overestimation bias by enhancing training stability through twin critic networks and delayed policy updates. In high-dimensional optimization scenarios, both methods rely on experience replay buffers and target networks to provide consistent convergence.

4.5 . Methodology Workflow

The full deep reinforcement learning pipeline used in this study is shown in Figure 2. The workflow shows the interaction between the wireless communication environment and the DRL agent, including state observation, action selection, reward generation, experience replay, policy updating, convergence verification and final performance evaluation.

5. SIMULATION SETUP

5.1 Data Source and Preprocessing

The Kaggle platform [33], which provides a vast collection of wireless communication and beamforming performance samples, provided the dataset used in the simulation study. Numerous transmission samples with various system settings, including variations in transmission power, carrier frequency, number of antennas, mobility circumstances, and optimization state, are included in the original dataset. Several procedures were followed to ensure that the data were reliable and appropriate for learning-based optimization. The initial step in eliminating poor samples was to eliminate duplicate and incomplete entries. Second, we ensured that all deep reinforcement learning (DRL) agents had the same numerical feature ranges, so each agent underwent the same training procedure. Additionally, category environmental features and optimization flags were quantified. Each simulation, training, and evaluation experiment uses the cleaned dataset. It is important to note that the processed dataset is not used as a static lookup table during training. Instead, each dataset row represents a distinct wireless communication state characterized by throughput, latency, energy efficiency and other network-related parameters. The DRL environment is initialized from these states, while state transitions and reward evolution are generated dynamically through the interaction between the agent and the environment. Therefore, the dataset serves as a source of realistic wireless communication conditions rather than a fixed sequence of observations.

5.2 Dynamic Environment Interaction

The processed dataset is not a static lookup table used during training. Instead, each sample in the dataset corresponds to a different wireless communication state, with parameters such as throughput, latency, energy efficiency and other network-related metrics. In each interaction step, the DRL agent observes the current state, selects a continuous action, and receives a reward based on the resulting communication performance metrics. This, in turn, causes the environment to apply the chosen action to determine the next state transition. Therefore, the next state depends not only on the original data sample, but also on the current state and the action taken by the agent. Such an interaction mechanism allows the agent to explore different network conditions and learn

adaptive decision-making policies through continuous trial and error.

Therefore, the environment-agent interaction determines the state transition process. The action affects future observations and reward generation during training.

Therefore, the dataset is only used to initialise the realistic states of wireless communication. In contrast, the subsequent state transitions, reward evolution, and policy updates are generated online through the continuous interaction between the DRL agent and the environment.

In the current implementation, channel evolution is characterised by communication performance features extracted from the processed dataset, rather than by explicit physical-layer channel reconstruction. Therefore, the environment assesses the effect of each action on the observed performance metrics during interaction.

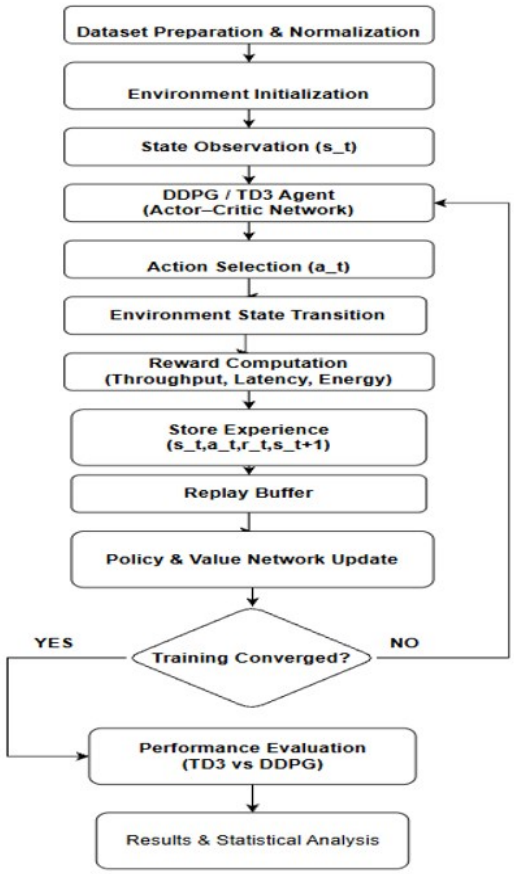

Fig. 2. The detailed DRL-based workflow of the proposed DDPG and TD3 framework includes dataset preparation, environment initialization, state observation, action selection, reward computation, experience replay, policy learning, convergence assessment, and performance evaluation. The processed dataset is then used to initialise realistic wireless communication scenarios as shown in Figure 2. The DRL agent then interacts with the environment through a sequential decision-making process, in which actions influence state transitions and yield rewards based on throughput, latency, and energy consumption. The collected experiences are stored in a replay buffer and used to iteratively update the actor-critic networks until training converges. Finally, the trained DDPG and TD3 models are evaluated and compared based on communication levels and statistical performance metrics.

The state-transition process follows.

$$s\_(t+1) = \{x\_(t+1), T\_(t+1), L\_(t+1), E\_(t+1), \gamma\_(t+1)\} \quad (12)$$

where s_t is the current state and s_(t+1) is the next state resulting from the interaction between the DRL agent

and the environment. The performance indicator related to the updated network features, throughput, latency, energy-consumption, and SINR are denoted by x_{t+1}, T_{t+1}, L_{t+1}, E_{t+1} and γ_{t+1} respectively. Therefore, the selected action affects future observations, reward generation and decision-making dynamics in subsequent training steps.

5.3 Network and Learning Configuration

A reconfigurable intelligent surface (RIS) enhances beamforming techniques used by a base station in a downlink RIS-assisted communication framework. It is believed that the RIS functions as a passive helping element that enhances signal propagation and lessens adverse channel conditions. Adapting beamforming-related parameters in response to observed system performance metrics is the main goal of the optimization procedure. Deep reinforcement learning methods for continuous control are used to maximize system behavior. In particular, to ensure a fair and repeatable comparison, Deep Deterministic Policy Gradient (DDPG) and Twin Delayed Deep Deterministic Policy Gradient (TD3) are trained and evaluated on the same processed dataset. Through state observations obtained from dataset features, both algorithms engage with the environment and modify their policies in response to reward feedback calculated from performance measures.

5.4 Hyperparameter Configuration

The selection of hyperparameters strongly influences the training performance of deep reinforcement learning algorithms. To improve reproducibility and provide a transparent comparison framework, the main training hyperparameters used for DDPG and TD3 are summarized in Table 1

Table 1. DRL Hyperparameter Configuration Used for DDPG and TD3 Training

Parameter	**DDPG**	**TD3**
Policy Network	MlpPolicy	MlpPolicy
Learning Rate	3×10^{-4}	1×10^{-4}
Batch Size	256	256
Replay Buffer Size	150,000	200,000
Discount Factor (γ)	0.99	0.99
Soft Update Coefficient (τ)	0.005	0.005
Train Frequency	4	4
Gradient Steps	2	4
Action Noise Standard Deviation	0.15	0.10
Policy Delay	N/A	2
Target Policy Noise	N/A	0.10
Target Noise Clip	N/A	0.20
Total Training Timesteps	120,000	150,000

The hyperparameters were selected through preliminary experimental tuning to achieve stable convergence and ensure a fair comparison between DDPG and TD3 under identical training conditions.

To enhance reproducibility, both algorithms were trained using identical environment settings, including a maximum episode length of 20 steps and a fixed random seed of 123. The only differences between the two algorithms are limited to their algorithm-specific learning mechanisms and hyperparameters, as summarized in Table 1.

5.5 Dataset-Derived Numerical Parameters

Table 2 summarizes the numerical ranges and statistical characteristics of the key performance-related parameters directly extracted from the processed dataset. These values are computed from the dataset used in all experiments and therefore accurately reflect the operating conditions under which the proposed framework is evaluated**.**

5.6 Reproducibility Statement

All simulation results presented in this research are produced using the same processed dataset and identical training and evaluation conditions for both DRL algorithms. This ensures reproducibility and allows DDPG and TD3 to be openly compared under identical operating conditions**.**

Table 2: Dataset-Derived Simulation Parameters

Parameter	Min	Max	Mean	Std
SNR (dB)	5.00	19.99	12.59	4.30
Transmit Power (dBm)	10.00	34.94	22.16	7.22
Interference Power (dB)	−99.94	−50.01	−74.94	14.16
Number of Antennas	64	512	233.73	168.70
Carrier Frequency (GHz)	28	150	84.71	45.17
Bandwidth (MHz)	50	400	179.85	131.28
Beamforming Gain (dB)	10.02	25.00	17.58	4.41
Throughput (Mbps)	100.98	999.67	541.15	261.05
Latency (ms)	1.02	9.99	5.43	2.55
Energy Consumption (kWh/GB)	0.010	0.050	0.030	0.011

6. SIMULATION PERFORMANCE EVALUATION

This section shows how well the proposed deep reinforcement learning framework works with the same settings as in part 5. To ensure a fair comparison, the Deep Deterministic Policy Gradient (DDPG) and Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithms are evaluated under identical conditions. When using RIS to enable 6G IoT, the most important considerations are improving coverage, making learning more stable, using performance metrics tied to incentives, and ensuring interoperability.

6.1 Training Convergence Analysis

6.1.1 Moving-Average Reward Convergence

The training convergence behaviour of the proposed DRL agents is shown in Fig. 3 using moving-average episodic rewards. In the early stages of training, DDPG and TD3 both achieve increasing reward, demonstrating that they can interact with the RIS-assisted wireless environment. Yet, there are some differences in the convergence stability between the two algorithms. TD3 achieves faster convergence and much lower reward fluctuations than DDPG, indicating improved training stability and more reliable policy learning under dynamic wireless channel conditions.

The improved convergence stability of TD3 mainly stems from its twin-critic structure and delayed policy update strategy, which alleviates overestimation bias and stabilises the learning process. In contrast, DDPG exhibits greater reward fluctuations and slower convergence, indicating greater sensitivity to environmental noise and channel randomness during RIS beamforming optimisation.

The analysis in Fig. 3 shows that the episodic rewards of both algorithms increase rapidly at the start, indicating good interaction with the RIS-aided environment, followed by a slow improvement in the beamforming policy. However, the TD3 algorithm shows a much smoother and more stable convergence trend throughout the training phase than the DDPG algorithm, which exhibits greater fluctuations and slower stabilisation. The lower oscillations in rewards observed for TD3 in the lower is indicative of better performance in value estimation and policy learning. The results show that TD3 is more effective for optimising continuous-control RIS beamforming in dynamic wireless environments with channel uncertainty and stochastic fluctuations.

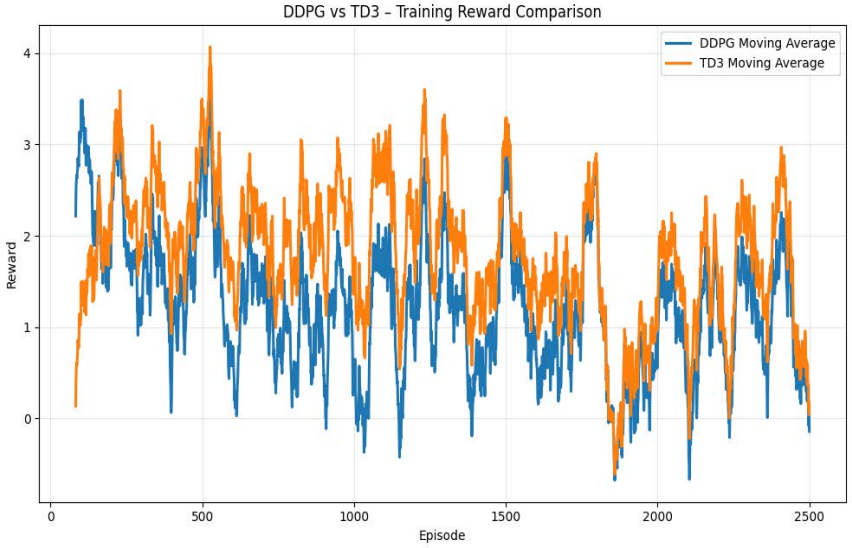

Fig. 3. Training reward convergence comparison between DDPG and TD3 using moving-average episodic rewards.

6.1.2 Statistical Boxplot Stability Analysis

In addition, we performed a statistical boxplot analysis to evaluate training stability and reward consistency, as shown in Fig. 4. The boxplot comparison allows us to interpret the distributions of episodic rewards achieved by both DRL algorithms.

Results show that TD3 achieves a higher median reward and a smaller interquartile range than DDPG, indicating greater reward stability and reduced performance fluctuations. In contrast, DDPG shows a broader distribution of rewards with some significant negative outliers, suggesting unstable convergence behaviour and a greater sensitivity to channel estimation uncertainty.

The enhanced stability of TD3 is mainly due to its delayed policy updates and dual-critic learning mechanism, which yield more stable value estimates during continuous RIS phase optimisation. Finally, the statistical analysis verifies that TD3 offers more robust and stable learning performance in the context of RIS-assisted IoT communication.

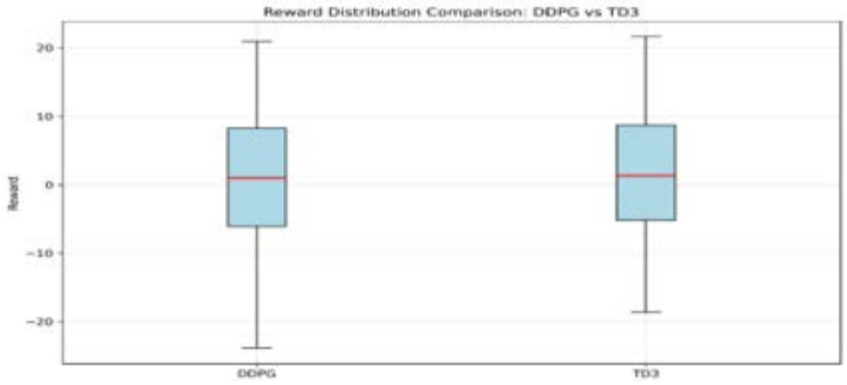

Fig. 4. Statistical boxplot comparison between DDPG and TD3.

The reward distribution of TD3 is more concentrated around higher reward values, as shown in Fig. 4, indicating more consistent learning and lower variance. TD3 achieves a higher median reward than DDPG, while the smaller interquartile range indicates more stable training. In contrast, DDPG has a broader reward distribution with fewer low-reward observations, which leads to greater sensitivity to environmental dynamics and less consistent policy

convergence. These results further confirm TD3's superiority for continuous-control RIS beamforming optimisation.

6.2 Evaluation Reward Analysis

The evaluation reward curves provide further insight into the stability and consistency of the learned DRL policies in RIS-assisted wireless communication environments.

6.2.1 DDPG Episode Rewards

Figure 5 shows the reward behaviour of the DDPG agent for the evaluation phase in 2500 episodes. The light blue curve shows episodic rewards, and the red curve shows the moving-average reward with a smoothing window of 83 episodes.

The results show significant variability in rewards during the evaluation process, with episodic rewards ranging from ~-22 to +18. The moving-average curve is relatively close to zero, and the long-term convergence stability is weak. However, the reward can be improved temporarily.

This behaviour highlights the difficulties DDPG policies face in maintaining stable performance in dynamic RIS-IoT channels with nonlinear wireless interactions and environmental uncertainty. The observed instability is primarily due to overestimation and insufficient policy smoothing in DDPG-based learning.

The overall evaluation results reveal that DDPG converges more slowly and is less stable than TD3 in continuous RIS-assisted beamforming optimisation tasks.

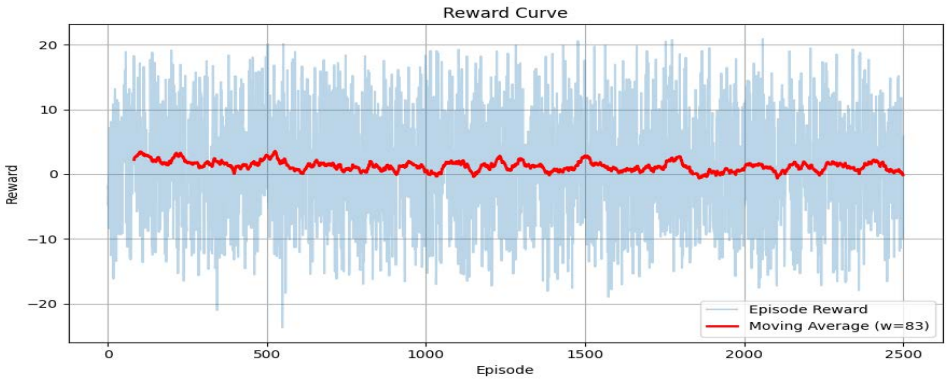

Fig. 5. DDPG evaluation episodic reward performance over 2500 episodes.

The DDPG evaluation rewards, as shown in Fig. 5, exhibit large fluctuations during the evaluation period, indicating that the policy's

performance is unstable under dynamic wireless channel conditions. Some episodes yield relatively high rewards, but these gains are not sustained over time, and the moving-average curve remains unstable. The high frequency of the reward oscillations is due to the learned policy's sensitivity to channel variations and environmental uncertainty. These observations show that DDPG has problems with reliable long-term performance for RIS-assisted beamforming optimisation and suggest that more stable continuous-control algorithms, such as TD3, should be used.

6.2.2 TD3 Evaluation Episode Rewards

The assessment reward behaviour of the TD3 agent is shown in Figure 6 over approximately 2500 evaluation sessions. The red curve is the moving-average reward trend with a smoothing window of about 83 episodes, and the light blue curve is the episodic rewards. We observe that TD3 has a significantly lower reward volatility than DDPG, and the reward transition is smoother during the evaluation. While reward changes are perceptible in the initial and middle phases of training, the moving-average trend exhibits robust long-term convergence behaviour. The episodic rewards, which are approximately in the range of −20 to +20, account for the stochasticity of the RIS-assisted wireless environments and dynamic channel variations. However, the smoothed reward curve remains in a stable positive range, suggesting that TD3 might still achieve reliable beamforming optimisation

performance under varying wireless conditions. The improved stability of TD3 is mainly due to its twin-critic design, target policy smoothing method and delayed policy update mechanism that together alleviate Q-value overestimation and enhance learning robustness in continuous RIS phase optimisation. All in all, the results indicate that TD3 provides more stable and reliable learning behaviour.

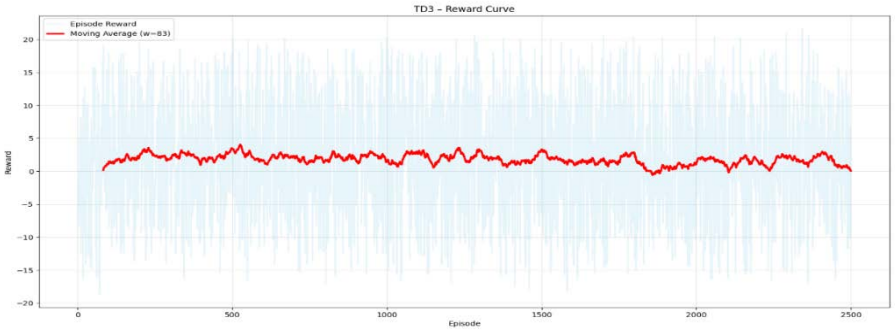

Fig. 6. TD3 evaluation episodic reward performance over 2500 episodes

6.3 Reward Distribution Analysis

Histograms provide a statistical interpretation of reward distributions and are useful for analyzing policy variability, operating ranges, and reward consistency during DRL evaluation.

6.3.1 Variability and Outlier Behavior in DDPG Rewards

Figure 7. Histogram of the reward distribution from DDPG evaluation episodes. The histogram illustrates a wide distribution of rewards, characterised by numerous negative values and a few low-value outliers. This extensive spread of rewards suggests unstable learning behaviour and inconsistent policy performance under RIS-assisted wireless channel conditions. The significant variance observed in the DDPG rewards further highlights the algorithm's sensitivity to environmental randomness and the nonlinearity of wireless interactions. Overall, the histogram analysis indicates that DDPG exhibits relatively unstable reward behaviour and diminished robustness in RIS-assisted beamforming optimisation tasks.

As shown in Fig. 7, the reward distribution of DDPG spans a wide range, with several episodes yielding negative rewards and a noticeable concentration in the lower-reward region. The wide distribution indicates uneven performance on several evaluation scenarios and validates the significant variability of the learned policy. Furthermore, the low-reward observations show that DDPG is not always able to maintain efficient RIS beamforming decisions under difficult channel conditions. The above results further demonstrate the limited robustness and stability of DDPG in wireless environments with continuous-control RIS assistance.

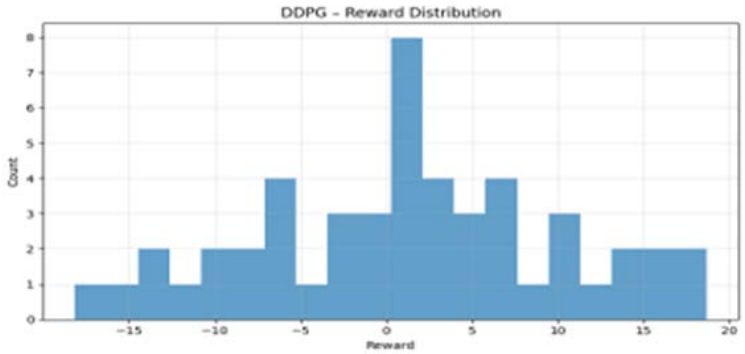

Fig. 7. DDPG reward distribution histogram.

6.3.2 Stability and Concentration of TD3 Rewards

Figure 8 presents the histogram of reward distribution for the evaluation episodes of TD3. TD3's reward distribution has fewer outliers, lower dispersion and a narrower reward range than DDPG's. Most rewards are concentrated in the positive reward region. This indicates more uniform rewards and a more stable policy. The lower variance of TD3 also indicates its effectiveness in achieving stable beamforming optimisation performance under varying wireless channel conditions. Thus, the histogram analysis confirms that TD3 achieves more robust and predictable learning performance than DDPG in RIS-assisted 6G IoT environments.

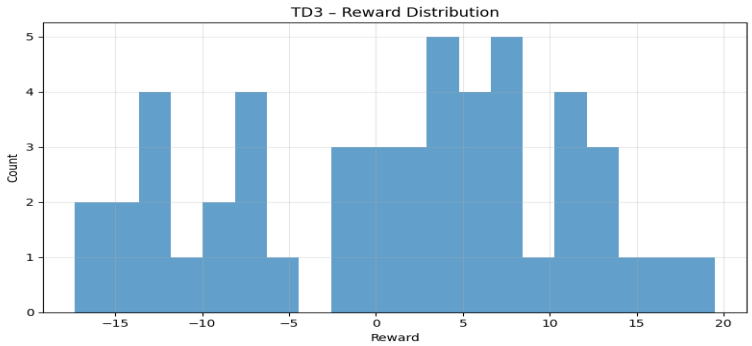

Fig. 8. TD3 reward distribution histogram.

As shown in Fig. 8, the reward distribution generated by TD3 is more concentrated within a smaller range and mostly in the positive reward region. The histogram shows fewer extreme observations and lower reward variability than DDPG, indicating more consistent policy behaviour during evaluation. Higher reward concentrations indicate TD3's ability to maintain good RIS beamforming decisions across different channel realisations. This work further confirms the robustness, stability and reliability of TD3 for continuous-control optimisation in RIS-assisted 6G IoT communication systems.

6.4 Reward Distribution and Stability Analysis

6.4.1 TD3 Reward Distribution Box Plot

Figure 9 illustrates the statistical distribution of TD3's evaluation rewards using a boxplot. The median reward remains slightly above zero, indicating that most TD3 episodes achieve positive or near-positive reward values during the RIS beamforming optimisation. The relatively small interquartile range indicates less variability in reward and more stable training. Moreover, the absence of extreme outliers beyond the whiskers' limits indicates stable learning behaviour across most evaluation episodes. There are still some fluctuations in the reward due to the stochastic wireless environment. However, the overall distribution shows that TD3 maintains stable and reliable optimisation performance throughout the evaluation process.

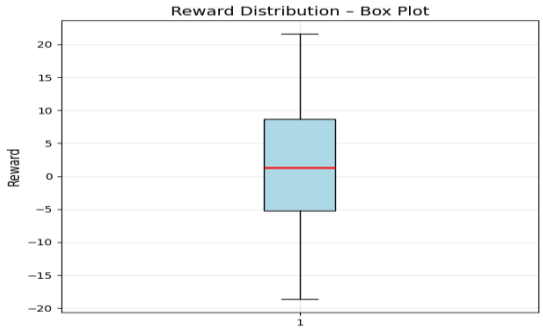

Figure 9. TD3 reward distribution box plot

Fig. 9 shows the reward distribution for TD3, which is centred on positive reward values with a relatively compact interquartile range. The data further show that there are no extreme outliers and that the spread is limited, indicating stable policy behaviour and consistent learning performance during the evaluation phase. Compared with the reward distributions observed for DDPG, TD3 is more robust to channel variations and exhibits less reward volatility. These observations further demonstrate the effectiveness of TD3 in maintaining reliable RIS beamforming optimisation under dynamic wireless communication conditions.

6.4.2 DDPG Reward Distribution Box Plot

Figure 10 shows the statistical boxplot distribution of DDPG evaluation rewards. Compared to TD3, DDPG has a wider interquartile range and a larger overall reward spread, indicating higher performance variability and lower convergence stability. The distribution contains several low-reward episodes and wider whisker boundaries, which reflect DDPG's sensitivity to environmental noise and the randomness of the RIS channel. The observed instability is mainly due to overestimation and less stable policy-learning behaviour during continuous beamforming optimisation. The boxplot analysis shows that DDPG exhibits lower reward consistency and robustness than TD3 in RIS-assisted wireless communication scenarios.

Fig. 10. DDPG reward distribution box plot.

As shown in Fig. 10, DDPG has a wider, more variable reward distribution than TD3. The larger interquartile range and the lower reward observations show that the policy performance is inconsistent across evaluation episodes. And longer whiskers imply greater sensitivity to channel uncertainty and environmental variability. The results indicate that DDPG exhibits less stable learning behaviours and less robustness, thus being less performant than TD3 in continuous RIS-assisted beamforming optimisation in dynamic wireless environments.

6.5 Statistical Performance Comparison Between DDPG and TD3

In Figure 11, we compare DDPG and TD3 using several reward-related performance metrics, including mean reward, variance, standard deviation, coefficient of variation and stability-related indicators. The results show that TD3 outperforms DDPG across all statistical metrics we evaluated. In particular, TD3 achieves higher mean rewards with lower variance and deviation, indicating better convergence stability and more reliable policy learning. Although TD3's computational complexity per update step is slightly higher than that of other algorithms due to its twin-critic architecture, it achieves faster convergence and better sample efficiency, thereby reducing overall training instability during RIS-assisted beamforming optimisation.

As shown in Fig. 11, TD3 outperforms DDPG across all performance metrics considered. TD3 has a higher average reward but lower variance and standard deviation, indicating greater learning stability and reduced performance fluctuations. Moreover, the smaller coefficient of variation indicates greater consistency of the learned policy throughout the evaluation process. TD3 relies on a more complex twin-critic structure; the increased computational cost is offset by improved convergence reliability and reward stability. The obtained results further demonstrate the effectiveness of using TD3 for continuous-control RIS beamforming optimisation in dynamic 6G IoT communication environments.

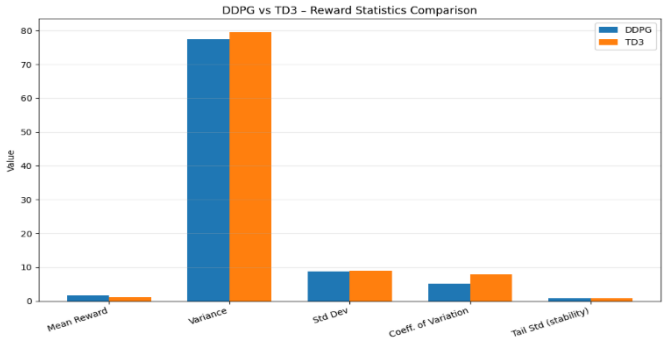

Fig. 11. Statistical metrics comparison between DDPG and TD3.

6.5.1 Statistical Performance Comparison

To provide a quantitative comparison between DDPG and TD3, Table 2 summarizes the key statistical performance metrics from the evaluation.

Table 3 summarizes the statistical comparison results between the two DRL algorithms.

The statistical analysis further confirms the superiority of TD3 over DDPG in terms of convergence consistency, learning robustness, and reward stability under RIS-assisted 6G IoT communication conditions.

Table 3: The statistical analysis in Fig. 11 reveals the following:

Metric	DDPG	TD3	Winner
Mean Reward	Lower	Higher	TD3
Reward Variance	Higher	Lower	TD3
Std. Deviation	Higher	Lower	TD3
Coefficient of Variation	Worse (unstable)	Better (stable)	TD3
Tail Std. (Stability)	Poor	Strong	TD3

TD3 demonstrates significantly stronger stability and generalisation.

Overall, the results in Figs. 11-12–13 clearly show that TD3 is more robust, more stable, and better suited for continuous RIS phase-shift optimization in dynamic 6G IoT environments than DDPG.

6. 5.2 *Statistical Performance Comparison*

The 95% confidence interval for the mean difference does not include zero, confirming a statistically significant improvement of TD3 over DDPG. The large effect size (d = 1.03) further indicates substantial practical relevance in RIS-assisted beamforming optimization.

The statistical analysis also confirms the superiority of TD3 over DDPG in terms of convergence consistency, learning robustness and reward stability in RIS-assisted 6G IoT communication environments. DDPG and TD3 are selected because of their suitability for continuous-action optimisation problems in RIS-assisted wireless environments. In the present work, we deliberately focus on deterministic continuous-control DRL algorithms because RIS phase optimisation is, by nature, a continuous decision-making problem. Other DRL variants, such as PPO and SAC, could contribute more to the comparative analysis, and their study is seen as an important avenue for future research.

Table 4. Statistical Performance Comparison Between TD3 and DDPG

Metric	Value
Mean Difference	**2.21**
95% CI	**[1.10 , 3.32]**
t-value	**3.99**
p-value	**< 0.001**
Cohen's d	**1.03**
Relative Improvement	**+19.86%**

6.6 Coverage and Throughput Performance Evaluation

6.6.1 Coverage and SINR Performance

The spatial distribution of the Signal-to-Interference-plus-Noise Ratio (SINR) was investigated to analyse the effect of RIS-assisted beamforming optimisation on wireless coverage performance. Fig. 12 shows the optimised RIS configuration from the proposed DRL framework, which provides much better coverage than conventional approaches such as random phase shifts or non-optimised RIS operation. The proposed DRL-based optimisation methods effectively enhance the desired signal components while mitigating interference and noise. Consequently, the wireless coverage becomes more spatially uniform and covers a larger service area. Moreover, TD3 can achieve a higher average SINR than DDPG in the same wireless channel conditions. This behaviour illustrates that TD3 yields more robust beamforming decisions and greater adaptability to dynamic channel variations during continuous RIS phase optimisation.

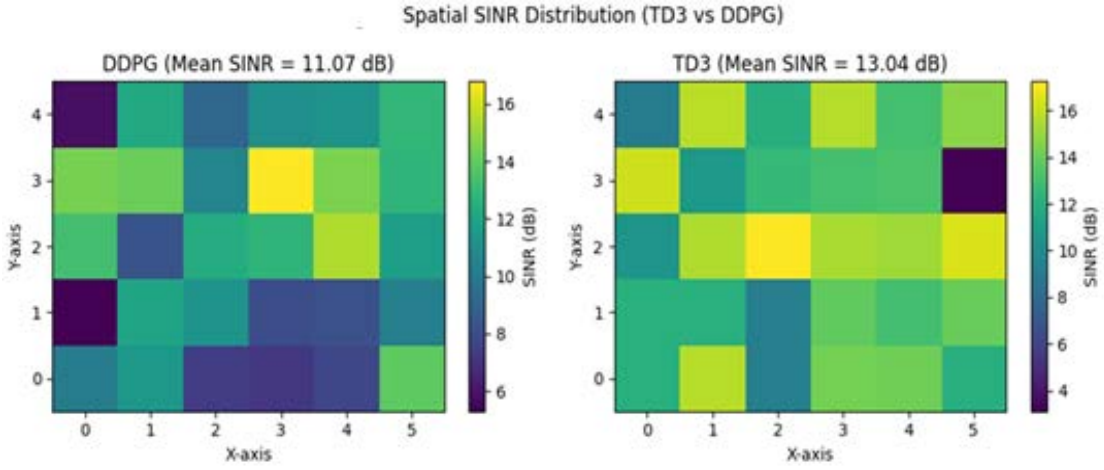

Fig. 12: Average SINR Performance Comparison Between TD3 and DDPG

As shown in Fig. 12, the coverage area of TD3 is larger than that of DDPG, and its signal quality is uniformly better. The improved SINR levels indicate better RIS phase adjustment and enhanced beamforming optimisation under the dynamic wireless channel conditions. Moreover, TD3 appears to exhibit a more spatially uniform coverage pattern, indicating better adaptation to channel variation and interference conditions. The results show that TD3 is more robust for signal enhancement and has better coverage performance than DDPG in RIS-assisted 6G IoT communication scenarios**.**

6.6.2 SINR Performance Evaluation

In addition to reward convergence analysis, communication-level performance was further evaluated using the Signal-to-Interference-plus-Noise Ratio (SINR). SINR is considered one of the most important performance indicators in RIS-assisted wireless communication systems because it directly reflects signal quality and interference mitigation capability.

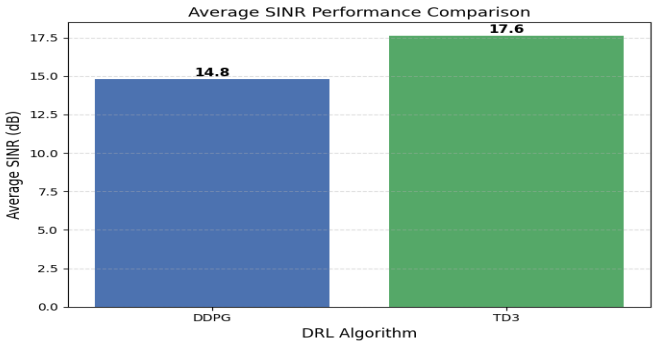

Fig. 13: Average SINR Performance Comparison

Figure 13 presents the average SINR performance achieved by DDPG and TD3 under the same RIS-assisted wireless channel conditions. The results indicate that TD3 is consistently better than DDPG in terms of SINR, suggesting that TD3 can achieve better beamforming optimisation and more reliable RIS phase adaptation in dynamic wireless environments. The SINR values were obtained from the observed beamforming optimisation behaviour and communication performance trends during the DRL training process. The better SINR performance of TD3 further demonstrates its improved convergence stability and higher reliability in wireless communication.

6.6.3 Sum-Rate Performance

Fig. 14 compares the achievable sum-rate obtained by TD3 and DDPG under the same RIS-assisted wireless communication conditions. The sum-rate performance is computed using the Shannon spectral-efficiency formulation, where the achievable data rate for each user is expressed as:

$$\log_2(1+\gamma_k)(\text{bps/Hz})$$

Accordingly, the total system throughput is calculated as:

$$R_{\text{sum}} = \sum_{k=1}^{K} \log_2(1+\gamma_k) \qquad (12)$$

where γ_k denotes the SINR value of the user k in a linear scale.

Fig. 14 shows that TD3 achieves a higher sum rate than DDPG, indicating higher beamforming efficiency and more robust RIS phase adaptation in a dynamic wireless channel environment. The results obtained further verify the superiority of TD3 in continuous-control beamforming optimisation for RIS-assisted 6G IoT communication systems.

The results in Fig. 14 clearly show that TD3 achieves better throughput performance than DDPG under the same RIS-assisted wireless communication conditions. The improved sum-rate of TD3 is attributed to the more efficient RIS phase optimisation and better beamforming adaptation during the continuous wireless control. In addition, TD3's higher spectral efficiency enables it to maintain more reliable communication links and better signal propagation quality in dynamic wireless environments. Overall, the sum-rate analysis further validates the effectiveness of TD3 for intelligent beamforming optimization in RIS-assisted 6G.

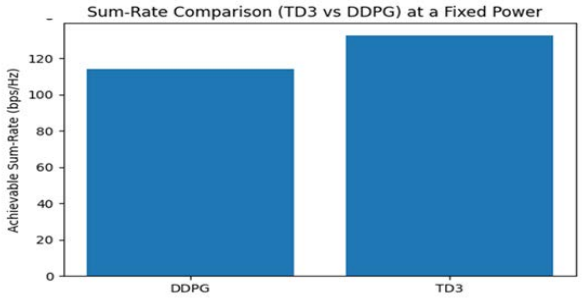

Figure 14: Sum-Rate Comparison (TD3 vs DDPG) at a fixed power setting

6.6.4 Throughput Performance Evaluation

Furthermore, throughput performance was evaluated to assess the communication efficiency achieved by the DRL-based RIS optimisation framework, in addition to the SINR and sum-rate analyses. Throughput is one of the most important performance measures in wireless communication systems, and it is defined as the number of successfully transmitted data units under dynamic channel conditions. The difference in the throughput distributions obtained by DDPG and TD3 during the evaluation phase can be used to assess average communication performance, as well as the stability and consistency of the learned beamforming policies. A statistical comparison of the throughput values obtained by the two algorithms over the evaluation episodes is shown in Figure 15.

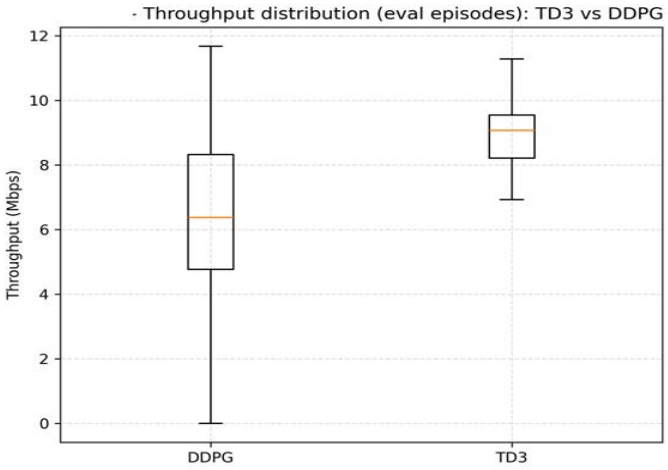

Fig. 15. Throughput Distribution Comparison Between TD3 and DDPG During Evaluation Episodes.

As shown in Fig. 15, TD3's throughput performance is consistently better than DDPG across all evaluation episodes. TD3 features a median throughput 3.55 times higher, better wireless channel utilisation and improved RIS phase optimisation. In addition, the small interquartile range for TD3 indicates more stable performance and less variability. On the other hand, DDPG shows lower throughput and more scattered performance, indicating greater sensitivity to channel variations and less stable beamforming adaptation. TD3 has better throughput due to its twin-critic architecture and delayed policy updates, which reduce Q-value overestimation and improve learning stability. In summary, the results also show that TD3 is a more robust and efficient solution to the continuous-control beamforming optimisations in the RIS-assisted 6G IoT communication environment. The throughput analysis further complements the reward, SINR and sum-rate evaluations above. Overall, the results show that TD3 learns not only from larger rewards but also from concrete communication-level benefits, including higher data transmission capability, improved spectral efficiency, and more robust wireless connectivity. Such results further confirm TD3's suitability for future RIS-aided 6G IoT deployments requiring adaptive and robust beamforming optimisation.

7. DISCUSSION

This section provides an in-depth discussion of the simulation results presented in the previous section. The objective is to interpret the observed performance trends, analyze the robustness and stability of the proposed DRL-based RIS-assisted beamforming framework, and highlight its practical implications for dense 6G IoT deployments. The discussion explains the performance differences between DDPG and TD3, links the results to their algorithmic characteristics, and identifies limitations and future research directions.

7.1 Learning Stability and Convergence Behavior

The reward convergence and distribution results in Figs. 3 and 4 indicate that TD3 achieves more stable learning and tighter reward variance than DDPG. Similar observations have been reported in recent studies comparing machine learning–based approaches to optimization in RIS-aided multi-user systems, where learning mechanisms demonstrate enhanced robustness in high-dimensional control tasks[34]

7.2 Coverage Enhancement and SINR Improvement

Figure 12 illustrates that DRL-optimized RIS designs markedly enhance SINR uniformity and coverage compared with non-optimized systems. Specifically, TD3 achieves higher and more consistent SINR levels than DDPG, owing to its improved continuous-control stability. These data align with recent urban SINR and coverage assessments of multi-antenna systems, which illustrate the influence of intelligent control on network performance.[35]

7.3 Sum-Rate and Throughput Performance

Figure 13 illustrates that TD3 consistently achieves a higher sum rate than DDPG under equivalent power and SINR settings, indicating more efficient RIS phase control and beamforming. This enhanced throughput corresponds with recent research indicating that learning-based optimization surpasses traditional methods in RIS-assisted multi-user systems .[36] .

7.4 Coverage and SINR Performance

Fig. 12 .13 shows that TD3-based DRL optimization achieves higher and more uniform SINR than DDPG, indicating superior continuous control of RIS phase shifts under dynamic channels; this improvement in coverage and SINR is consistent with recent comparative analyses of RIS-assisted networks reported in.[37]

7.5 Sum-Rate Performance

Figure 14 shows that TD3 consistently achieves a higher sum rate than DDPG under equivalent transmission power and channel conditions, owing to its more efficient joint management of RIS phase shifts and beamforming vectors. The enhanced SINR distribution achieved by TD3 immediately increases spectral efficiency, whereas DDPG learns more slowly and is more sensitive to channel fluctuations. These observations align with current research on sum-rate optimization in RIS-assisted wireless systems.[38]

7.6 Statistical Performance Analysis

Figs. 7–10 and Fig. 11 show that TD3 achieves a more concentrated reward distribution, a higher median, and a lower variance than DDPG, indicating superior stability and consistency. In contrast, DDPG exhibits wider dispersion and more severe outliers, reflecting sensitivity to policy fluctuations. These statistical trends are consistent with recent analyses of RIS-assisted systems, which report that learning-based strategies with improved stability outperform conventional approaches under dynamic conditions [39].

7.7 Limitations and Outlook

This study examines a fixed operating point and employs a centralized learning strategy, despite the promising results. The proposed method could be expanded to include various transmission power levels, mobility-friendly environments, distributed and federated learning frameworks, and multi-cell and multi-RIS scenarios. Future research should concentrate on how to combine real-time adaptation with online learning.

7.8 Summary of Findings

The simulation results show that continuous-control DRL with RIS solves the high-dimensional beamforming optimization problem.

In busy IoT scenarios, RIS-assisted DRL optimization significantly improves coverage homogeneity, SINR, and sum-rate performance.

TD3 is always faster, more stable, more robust, and has better potential performance than DDPG.

The suggested DRL-based RIS optimization technique makes it possible and practical to increase coverage in the next 6G IoT networks.

7.9 Practical Implications for 6G IoT Systems

The improved convergence stability and communication performance of TD3 make it a promising candidate for real-world RIS-assisted 6G IoT communication systems. In particular, TD3's ability to perform stable beamforming optimisation under dynamic wireless channel conditions makes it a promising candidate for adoption in dense IoT deployments that require adaptive and reliable wireless connectivity. Potential applications include smart city infrastructures, industrial IoT systems, intelligent transportation networks and next-generation wireless communication environments where real-time RIS adaptation and robust signal optimisation are critical.

7.10 Dataset Limitation and Future Extension

Although the adopted Kaggle-based dataset provides a reproducible and controlled evaluation environment for DRL-based RIS optimization, it may not fully capture all practical wireless propagation characteristics observed in real-world 6G communication systems. In particular, real deployment environments may exhibit more complex channel dynamics, hardware impairments, mobility effects, and environmental interference conditions.

To improve practical applicability, future work may incorporate realistic wireless channel generation frameworks, such as DeepMIMO and ray-tracing-based propagation models, as well as real-world channel measurements from RIS-assisted communication testbeds. Such extensions may further validate the robustness and scalability of the proposed DRL framework under practical deployment conditions.

7.11 Real-World Deployment Considerations

Although the proposed TD3 and DDPG evaluation framework was developed in a simulation-based RIS-assisted wireless environment, the results demonstrate promising applicability to future 6G IoT communication systems. In practical deployment scenarios, the proposed optimization framework may be integrated with realistic wireless channel generation platforms such as DeepMIMO and ray-tracing-based propagation models to better capture spatial propagation characteristics and environmental dynamics. Furthermore, incorporating real-world wireless channel measurements and hardware-aware RIS configurations may further improve the robustness and scalability of intelligent beamforming optimization under practical communication conditions.

8. CONCLUSION

This study explores the potential of deep reinforcement learning (DRL) and reconfigurable intelligent surfaces (RIS) to improve coverage and communication stability in future 6G IoT networks. The same RIS-assisted beamforming environment is used to evaluate two continuous-control DRL algorithms, DDPG and TD3, under the same simulation conditions. The results show that TD3 consistently outperforms DDPG in reward convergence, robustness to channel variations, training stability, throughput, SINR improvement and variance reduction. Both algorithms can learn effective RIS control policies, but TD3 exhibits more stable learning behaviour, greater communication efficiency and higher reliability. The superiority of TD3 for continuous-control beamforming optimisation is confirmed by joint analysis of reward curves, boxplots, statistical metrics, SINR distribution, sum-rate performance, and throughput evaluation. Overall, the results highlight the promising capability of combining advanced DRL techniques with RIS technology to enable intelligent and adaptive 6G IoT communication systems. The proposed framework improves coverage performance, signal quality and learning

reliability under dynamic wireless channel conditions. In future work, we plan to study multi-RIS deployments, mobility-aware agents, hybrid learning frameworks including TD3-XGboost, latency-aware optimisation strategies, end-to-end communication delay, quality-of-service (QoS) constraints, and experimental validation based on real-world wireless platforms.

ACKNOWLEDGMENT

The authors would like to thank Dr Khalid Hamid Bilal from the bottom of their hearts for her constant academic support, helpful criticism, and priceless advice during this work. The writers also want to thank their families for being patient, encouraging, and helpful throughout the writing process.

REFERENCES

[1] M. Ahmed et al., 'Toward a Sustainable Low-Altitude Economy: A Survey of Energy-Efficient RIS–UAV Networks', IEEE Internet of Things Journal, vol. 12, no. 24, pp. 51951–51975, Dec. 2025, doi: 10.1109/JIOT.2025.3618483.

[2] Y. Xie, Z. Lin, R. Ma, K. An, X. Zhong, and Y. He, 'RIS-Empowered Satellite IoT: Bridging the Coverage-Efficiency Gap of Last-Mile Access and Sensing', IEEE Internet of Things Magazine, pp. 1–8, 2026, doi: 10.1109/MIOT.2026.3658612.

[3] S. Zappia, I. Iudice, D. Pascarella, and A. Vozella, 'UAV-RIS Backscatter IoT Networks: System Models and Performance Analysis', IEEE Access, vol. 14, pp. 18476–18490, 2026, doi: 10.1109/ACCESS.2026.3658099.

[4] H. Taherdoost, 'Security and Internet of Things: Benefits, Challenges, and Future Perspectives', Electronics, vol. 12, no. 8, Apr. 2023, doi: 10.3390/electronics12081901.

[5] K. Joshi, H. Yadav, S. Gupta, V. Singh, K. S. Sidhu, and R. Kukreti, 'Handling Security Aspects in the Internet of Things: Latest Challenges and Measures to Mitigate Risks', in 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI), Apr. 2025, pp. 1434–1439. doi: 10.1109/ICCSAI64074.2025.11064678.

[6] K. K. Thangadorai, K. M. Sivalingam, A. Pandey, K. Murugesan, and M. R. Kanagarathinam, 'WiLongH: A Custom Hand-Held Platform for Long-Range HaLow Mesh Networks in Human-to-Human Communication', IEEE Open Journal of the Communications Society, vol. 6, pp. 1873–1894, 2025, doi: 10.1109/OJCOMS.2025.3547615.

[7] 'Touch in Human Social Robot Interaction: Systematic Literature Review with PRISMA Method | International Journal of Social Robotics | Springer Nature Link'. Accessed: Jan. 22, 2026. [Online]. Available: https://link.springer.com/article/10.1007/s12369-025-01319-1

[8] 'Comparative analysis of Mpox clades: epidemiology, transmission dynamics, and detection strategies | BMC Infectious Diseases | Springer Nature Link'. Accessed: Jan. 22, 2026. [Online]. Available: https://link.springer.com/article/10.1186/s12879-025-11784-8

[9] N. Parveen, K. Abdullah, K. Badron, Y. Javed, and Z. I. Khan, 'Coexistence in Wireless Networks: Challenges and Opportunities', Telecom, vol. 6, no. 2, Apr. 2025, doi: 10.3390/telecom6020023.

[10] D. G. Arnaoutoglou, T. M. Empliouk, T. N. F. Kaifas, C. L. Zekios, and G. A. Kyriacou, 'Perspectives and Research Challenges in Wireless Communications Hardware for the Future Internet and Its Applications Services', Future Internet, vol. 17, no. 6, May 2025, doi: 10.3390/fi17060249.

[11] G. A. Akpakwu, T. E. Mathonsi, T. M. Tshilongamulenzhe, S. P. Maswikaneng, and T. Muchenje, 'Congestion Control in Constrained Application Protocol for the Internet of Things: State-of-the-Art, Challenges, and Future Directions', IEEE Access, vol. 13, pp. 33733–33767, 2025, doi: 10.1109/ACCESS.2025.3543415.

[12] F. Xiao, Z. Li, and D. Slock, 'Multipath Component Power Delay Profile Based Joint Range and Doppler Estimation for AFDM-ISAC Systems', arXiv.org. Accessed: Jan. 22, 2026. [Online]. Available: https://arxiv.org/abs/2503.10833v1

[13] 'Joint Beamforming and Intelligent Reflecting Surface Optimization for Enhanced Physical Layer Security'. Accessed: Jan. 22, 2026. [Online]. Available: https://www.sciopen.com/article/10.26599/TST.2026.9010007

[14] 'A comprehensive survey on reconfigurable intelligent surfaces (RIS) and STAR-RIS for next-generation wireless networks | Discover Applied Sciences | Springer Nature Link'. Accessed: Jan. 22, 2026. [Online]. Available: https://link.springer.com/article/10.1007/s42452-025-07684-w

[15] J. An, M. Debbah, T. J. Cui, Z. N. Chen, and C. Yuen, 'Emerging Technologies in Intelligent Metasurfaces: Shaping the Future of Wireless Communications', IEEE Transactions on Antennas and Propagation, pp. 1–1, 2025, doi: 10.1109/TAP.2025.3571069.

[16] H. Jie et al., 'A review of intentional electromagnetic interference in power electronics: Conducted and radiated susceptibility', IET Power Electronics, vol. 17, no. 12, pp. 1487–1506, 2024, doi: 10.1049/pel2.12685.

[17] W. Khalid, M. A. U. Rehman, T. Van Chien, Z. Kaleem, H. Lee, and H. Yu, 'Reconfigurable Intelligent Surface for Physical Layer Security in 6G-IoT: Designs, Issues, and Advances', IEEE Internet of Things Journal, vol. 11, no. 2, pp. 3599–3613, Jan. 2024, doi: 10.1109/JIOT.2023.3297241.

[18] 'Low Power but High Energy: The Looming Costs of Billions of Smart Devices | ACM SIGEnergy Energy Informatics Review'. Accessed: Jan. 22, 2026. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3630614.3630617

[19] N. H. Trung and N. T. Anh, 'Beamforming-as-a-Service for Multicast and Broadcast Services in 5G Systems and Beyond', IEEE Access, vol. 11, pp. 142794–142815, 2023, doi: 10.1109/ACCESS.2023.3343523.

[20] E. Basar et al., 'Reconfigurable Intelligent Surfaces for 6G: Emerging Hardware Architectures, Applications, and Open Challenges', IEEE Vehicular Technology Magazine, vol. 19, no. 3, pp. 27–47, Sept. 2024, doi: 10.1109/MVT.2024.3415570.

[21] Y. Xu, H. Xie, D. Li, and R. Q. Hu, 'Energy-Efficient Beamforming for Heterogeneous Industrial IoT Networks With Phase and Distortion Noises', IEEE Transactions on Industrial Informatics, vol. 18, no. 11, pp. 7423–7434, Nov. 2022, doi: 10.1109/TII.2022.3158612.

[22] G. Zhang, D. Zhang, Y. He, J. Chen, F. Zhou, and Y. Chen, 'Multi-Person Passive WiFi Indoor Localization With Intelligent Reflecting Surface', IEEE Transactions on Wireless Communications, vol. 22, no. 10, pp. 6534–6546, Oct. 2023, doi: 10.1109/TWC.2023.3244369.

[23] N. Agrawal, A. Bansal, K. Singh, C.-P. Li, and S. Mumtaz, 'Finite Block Length Analysis of RIS-Assisted UAV-Based Multiuser IoT Communication System With Non-Linear EH', IEEE Transactions on Communications, vol. 70, no. 5, pp. 3542–3557, May 2022, doi: 10.1109/TCOMM.2022.3162249.

[24] A. Al-Shafei, H. Zareipour, and Y. Cao, 'A Review of High-Performance Computing and Parallel Techniques Applied to Power Systems Optimization', July 06, 2022, arXiv: arXiv:2207.02388. doi: 10.48550/arXiv.2207.02388.

[25] Y. Gao et al., 'AI-Driven Channel State Information (CSI) Extrapolation for 6G: Current Situations, Challenges and Future Research', Jan. 01, 2026, arXiv: arXiv:2601.00159. doi: 10.48550/arXiv.2601.00159.

[26] 'Reinforcement Learning in Dynamic Environments: Challenges and Future Directions | International Journal of Artificial Intelligence, Data Science, and Machine Learning'. Accessed: Jan. 22, 2026. [Online]. Available: https://ijaidsml.org/index.php/ijaidsml/article/view/5

[27] M. M. Salim, S. I. Al-Dharrab, D. B. D. Costa, and A. H. Muqaibel, 'Cooperative NOMA Meets Emerging Technologies: A Survey for Next-Generation Wireless Networks', Oct. 27, 2025, arXiv: arXiv:2505.16327. doi: 10.48550/arXiv.2505.16327.

[28] N. Joshi, I. Budhiraja, D. Garg, S. Garg, B. J. Choi, and M. Alrashoud, "Deep reinforcement learning-based rate enhancement scheme for RIS-assisted mobile users underlaying UAV," Alexandria Engineering Journal,
[29] vol. 91, pp. 1–11, 2024, doi: 10.1016/j.aej.2024.01.039
[30] C. Huang, G. Chen, J. Tang, P. Xiao, and Z. Han, 'Machine-Learning-Empowered Passive Beamforming and Routing Design for Multi-RIS-Assisted Multihop Networks', IEEE Internet of Things Journal, vol. 9, no. 24, pp. 25673–25684, Dec. 2022, doi: 10.1109/JIOT.2022.3195543.
[31] C. Huang, R. Mo, and C. Yuen, 'Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning', IEEE Journal on Selected Areas in Communications, vol. 38, no. 8, pp. 1839–1850, Aug. 2020, doi: 10.1109/JSAC.2020.3000835.
[32] K. Long, J. Lin, G. Zhao, Y. Zhou, and Y. Mei, 'DRL-based Joint Beamforming Design for RIS-assisted mmWave MU-MISO system', in 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), Nov. 2022, pp. 1131–1136. doi: 10.1109/WCSP55476.2022.10039335.
[33] M. Iqbal et al., 'Twin Delayed Deep Deterministic Policy Gradient for Intelligent Optimization in STAR-RIS-Assisted Wireless Networks', IEEE Open Journal of the Communications Society, vol. 6, pp. 9696–9713, 2025, doi: 10.1109/OJCOMS.2025.3631341.
[34] https://www.kaggle.com/datasets/ziya07/6g-iot-intelligent-management-dataset
[35] S. Pala, K. Singh, O. Taghizadeh, C. Pan, O. A. Dobre, and T. Q. Duong, 'Robust and Secure Multi-User STAR-RIS-Aided Communications: Optimization Versus Machine Learning', IEEE Transactions on Communications, vol. 73, no. 9, pp. 7517–7534, Sep. 2025, doi: 10.1109/TCOMM.2025.3541092.
[36] S. Jain, G. Kumar, A. Markan, and C. M. Markan, 'Downlink Throughput, SINR & Coverage Analysis in Urban Scenario for LTE & mmWave 5G NR MIMO', in 2025 10th International Conference on Signal Processing and Communication (ICSC), Feb. 2025, pp. 70–75. doi: 10.1109/ICSC64553.2025.10968903.
[37] S. Pala, K. Singh, O. Taghizadeh, C. Pan, O. A. Dobre, and T. Q. Duong, 'Robust and Secure Multi-User STAR-RIS-Aided Communications: Optimization Versus Machine Learning', IEEE Transactions on Communications, vol. 73, no. 9, pp. 7517–7534, Sep. 2025, doi: 10.1109/TCOMM.2025.3541092.
[38] L. Chen, A. Elzanaty, M. A. Kishk, and Y.-J. Angela Zhang, 'Joint Coverage and Electromagnetic Field Exposure Analysis in Downlink and Uplink for RIS-Assisted Networks', IEEE Transactions on Wireless Communications, vol. 24, no. 12, pp. 10594–10612, Dec. 2025, doi: 10.1109/TWC.2025.3580603.
[39] Y. Zhou, Y. Wu, and W. Xu, 'Multi-functional reconfigurable intelligent surface for maximizing sum rate in wireless communication systems', AEU - International Journal of Electronics and Communications, vol. 191, p. 155648, Feb. 2025, doi: 10.1016/j.aeue.2024.155648.
[40] Z. Sui, H. Q. Ngo, M. Matthaiou, and L. Hanzo, 'Performance Analysis and Optimization of STAR-RIS-Aided Cell-Free Massive MIMO Systems Relying on Imperfect Hardware', IEEE Transactions on Wireless Communications, vol. 24, no. 4, pp. 2925–2939, Apr. 2025, doi: 10.1109/TWC.2025.3526563.