SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I3P109

Volume 13 Issue 3, 119-128, March 2026

Original Article

Deep Reinforcement Learning for Spectral Efficiency in Terahertz-Enabled 6G RANs

Sheetal Vishal Deshmukh[1], Shahanawaj Ahamad[2], P S V Srinivasa Rao[3], G. Meena Devi[4]

[1]*Department of Computer Application, Bharati Vidyapeeth Yashwantrao Mohite Institute of Management, Karad, Maharashtra, India.*
[2]*Department of Software Engineering, College of Computer Science and Engineering, University of Hail, Hail, Saudi Arabia.*
[3]*Department of Computer Science and Engineering, Joginpally B R Engineering College, Telangana, India.*
[4]*Department of Mathematics, St. Joseph's College of Engineering, Tamil Nadu, India.*

[1]*Corresponding Author : sheetalvishaldeshmukh05@gmail.com*

Received: 08 December 2025 Revised: 10 January 2026 Accepted: 10 February 2026 Published: 23 March 2026

Abstract - *Terahertz communication is a fundamental component towards the realization of the highest possible data rate that is currently being conceived in the 6G Radio Access Networks. The bands of Terahertz frequencies have large path loss and significant molecular absorption and thus require accurate and dynamic resource management. To maximize Spectral Efficiency while balancing energy constraints, this research proposes a robust Deep Reinforcement Learning system that is built upon a Multi-Objective Double Deep Q-network framework, including the environment, sensing, intelligence, and performance layers built using the Deep MIMO ray-tracing dataset to create a high-fidelity digital twin of the Terahertz channel. Experiments show that the framework can overcome overestimation bias by performing extensive normalization, state-vectorizing, and multi-objective reward shaping to stabilize the learning process, which achieves a rapid convergence and higher stability than conventional Deep Q Network methods. The proposed model achieves significant improvements in Spectral Efficiency (24.1 bps/Hz), Energy Efficiency (4.8 Gbits/J), Throughput Satisfaction Rate (up to 95%), and sub-framework beam alignment latency (0.8 ms). The results show the effectiveness of the Deep Reinforcement Learning methodologies in solving complex propagation problems in the 6G Terahertz communication systems.*

Keywords - *Deep Reinforcement Learning, 6G Wireless Networks, Terahertz Communication, Radio Access Networks, Multi-Objective Double Deep Q-Network.*

1. Introduction

The Sixth-Generation (6G) wireless networks provide the rates of terabit per second as well as sub-milliseconds latency, which will support impactful applications of holography telepresence, immersive extended reality, and digital twins [1, 2]. One of the most promising enabling technologies for these aspirations is the Terahertz (THz) band (0.10-10 THz), which offers a bandwidth previously unheard of that is orders of magnitude higher than sub-6GHz and millimeter-wave bands [3]. The Terahertz communication has high potential; however, it faces serious challenges in the physical layer. The excessively large path loss, atmospheric molecular absorption frequency-dependent, extremely small beamwidths, and extreme loss to blockage and mobile users combine to create highly dynamic, non-stationary propagation conditions that may not be well represented by traditional resource allocation models [3]. As a result, Spectral Efficiency (SE) is prone to deterioration when operating within these constraints, and the corresponding latency of beam-alignment is actually staggering. A significant body of research literature has focused on the optimization of a single performance measure, usually SE or throughput, in millimeter-wave (mmWave) systems that use simplified stochastic channel models [4, 5].

The complexity of the interdependence of SE, Energy Efficiency (EE), and latency which are critical in a realistic THz operating environment. Additionally, the conventional Deep Q-Networks (DQN) have the disadvantage of omission of bias, which leads to unstable convergence when applied in highly dynamic environments of THz channels [6]. To overcome such shortcomings, this paper proposes a Multi-Objective Double Deep Q-Network (MO-DDQN) framework. Unlike prior single-objective Deep Reinforcement Learning (DRL) methods designed for mmWave systems [7], the proposed framework integrates site-specific DeepMIMO ray-tracing at 140 GHz, incorporates molecular absorption characteristics through the Beer-Lambert law directly into reward shaping, and employs Double DQN logic to mitigate overestimation bias. This enables balanced, stable optimization of SE, EE, and beam alignment latency in THz-enabled 6G Radio Access Networks.

The remainder of this paper is as follows. Section 2 presents a systematic literature review of the existing

research on deep reinforcement learning in wireless networks, by pointing out the main limitations involved with the application of these techniques to THz communications. Section 3 will contain an in-depth presentation of the THz channel model, the general system architecture, and the description of the proposed multi-objective framework. Section 4 presents the simulation setup, key results, and a detailed and rigorous comparison with baseline methods. The paper ends in Section 5 with a brief summary of findings and an outline of prospective avenues of future work.

2. Literature Review

DRL has become an effective tool of resource management in modern 5G and 6G network architectures. Early works successfully applied DQN and its variants to problems such as power control, subcarrier allocation, and beamforming in sub-6 GHz and mmWave systems [8, 9]. These studies demonstrated that DRL could adapt to dynamic traffic and interference conditions far better than traditional optimization techniques. Recently moved to work in higher frequency space, some studies have investigated how the techniques of DRL can be applied in the fields of THz beamforming and resource allocation. A framework, using deep Q -network (DQN) to solve a trajectory optimization problem in UAV -IRS -assisted THz networks, was introduced by Saleh et al. (2024), whereby substantial energy consumption and mission completion time reductions were realized. In a similar manner, Omar et al. (2023) used DRL to increase the capacity of systems with the help of flying IRS modules in terahertz applications [10]. Longer ago, Balaji et al. (2025) combined adaptive beamforming with DQN in RIS-assisted THz systems, with further improvement in performance.

Although this has been made, major gaps still remain. To start with, most of the studies fail to optimize multiple objectives; most of them simply maximize the sum rate or throughput and ignore the intra-objective conflicts between spectral efficiency, energy efficiency, and latency in THz bands [11]. Second, the majority of frameworks are based on simplified statistical models of channels that do not reflect any spatial consistency, effects of blockage, and frequency-selective molecular absorption, which are the leading effects in real THz propagation [12]. Third, classical DQN forms still can be affected by the overestimation bias, which causes unreliable learning and generalization in non-stationary THz settings. These gaps are summarized in Figure 1.

This paper aims to solve the current shortcomings by developing a Multi-Objective Double Deep Q-Network (MO-DDQN) that is implemented in a four-layer system. The proposed framework, through integrating high-fidelity DeepMIMO ray-tracing, physics-based reward-shaping scheme, and Double DQN decoupling, optimises Spectral Efficiency (SE), Energy Efficiency (EE), and beam alignment latency, and has a faster and more stable convergence than traditional deep reinforcement learning methods.

3. A Multi-Objective Terahertz-Enabled Deep Q-Network Framework for 6G RANs

To optimize the SE in the THz spectrum, an MO-DDQN framework is proposed. The proposed methodology is a combination of high-fidelity channel modeling and adaptive reinforcement learning mechanisms that is designed to effectively manage the highly variable channel dynamics inherent in 6G RANs.

3.1. Research Design and System Architecture

This research follows a closed-loop architectural framework that has four distinct functional layers, as shown in Figure 2, to ensure a continuous data throughput as well as low-latency decision making.

The Environment Layer serves as the physical substrate by creating site-specific channel realisations using DeepMIMO ray-tracing. This raw data then goes on to feed the Sensing Layer, which transforms stochastic channel state information into a vectorized set of features. The central Intelligence Layer applies an MO-DDQN to make a complex trade-off between throughput and energy consumption with decoupled target networks. Finally, the Performance Layer uses the ns-3 simulator[13] to rigorously evaluate the agent's policies against stringent 6G Key Performance Indicators (KPIs) to ensure that the learning agent achieves high throughput satisfaction and low alignment latency.

3.2 Data Collection and Pre-Processing

Of the spectrum[14], the input raw signal represented in Equation (1). The physical channel is mathematically formalised by the channel matrix H, which involves a setup of N antennas and L discrete propagation paths, as represented in Equation (2). By incorporating the ray-tracing information, the model captures the realistic spatial consistency and path loss for providing the deterministic basis for a reinforcement-learning agent to learn the optimal beam forming strategies in a high fidelity 6G RAN simulation. To characterize the ultra-wideband propagation environment, use the DeepMIMO dataset, in particular the 01 outdoor scenario in the carrier frequency of 140 GHz [15]. This high-frequency modelling represents the inherent sparsity and atmospheric absorption phenomena in the THz part.

$$y = Hx + n \tag{1}$$

$$\mathrm{H} = \sum_{l=1}^{L} \alpha_l \, \mathrm{a}_{\mathrm{rx}}(\theta_l, \phi_l) \mathrm{a}_{\mathrm{tx}}^{H}(\theta_l, \phi_l) e^{-j2\pi\tau_l f_c} \tag{2}$$

where x: Transmitted symbol vector after precoding and power allocation. H: Channel matrix which describes multipath propagation, attenuation, absorption, and DeepMIMO simulation parameters. n: Additive white Gaussian noise vector. y: Received raw signal vector after THz propagation. L denotes the number of dominant propagation paths, α_l is the complex gain of the l-th path (incorporating path loss, molecular absorption, and reflection losses), $\mathbf{a}_{\mathrm{rx}}$ and $\mathbf{a}_{\mathrm{tx}}^{H}$ are the receive and transmit

array response vectors for the corresponding azimuth θ_l and elevation ϕ_l angles of arrival/departure, fc=140 GHz is the carrier frequency, and τ_l represents the path delay. This formulation ensures spatially consistent, site-specific channel realizations critical for evaluating beamforming and resource allocation in highly directional THz systems.

Fig. 1 Literature gaps in DRL for THz-6G resource management

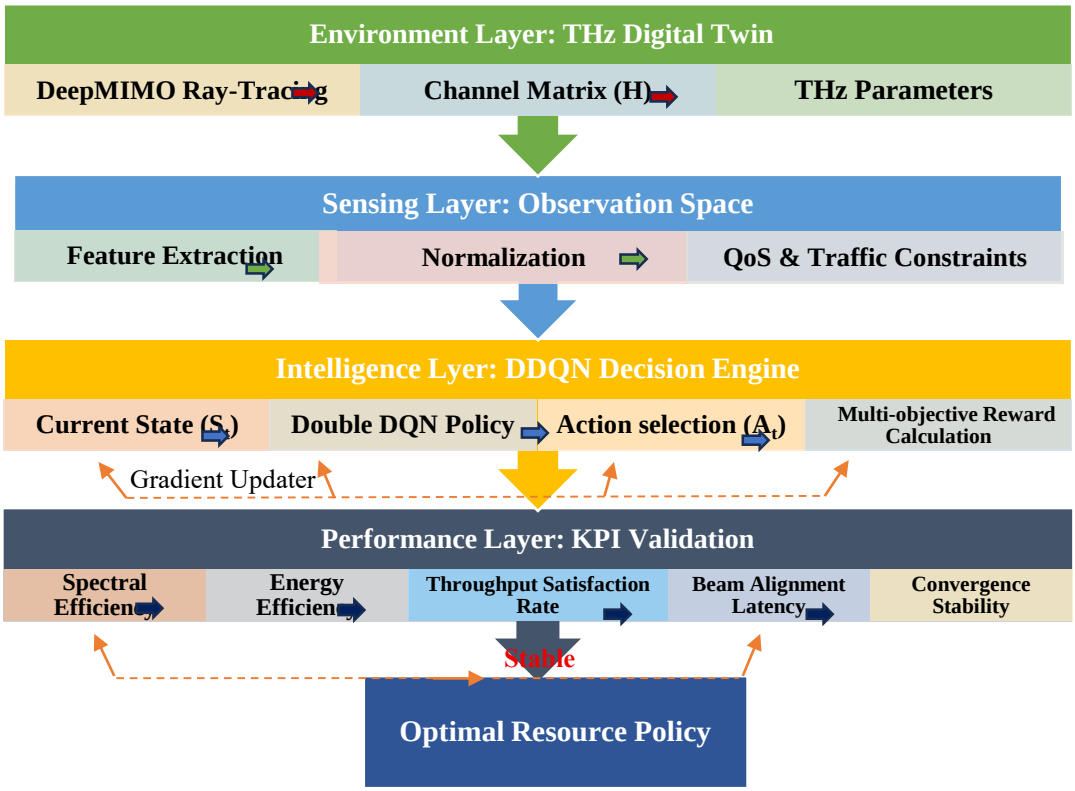

Fig. 2 Four-layer closed-loop DRL framework for THz-6G

To simulate the propagation impairments involved with the THz band, the molecular absorption loss is taken into account in the multipath channel model by means of the Beer-Lambert law [16]. The linear absorption factor $L_{\text{abs}}(f,d)$ for a particular frequency f and distance d is defined as follows in Equation (3).

$$L_{\text{abs}}(f,d) = e^{\kappa(f)\,d} \tag{3}$$

Where $\kappa(f)$(Np/m) is the frequency-dependent absorption coefficient. This factor modifies each path gain as $\alpha_l^{'} = \frac{\alpha_l}{\sqrt{L_{\text{abs}}(f_c,d_l)}}$($f_c$ = 140 GHz, d_l: path length), ensuring the channel matrix H in Equation (1) captures realistic gaseous absorption effects within the DeepMIMO framework.

To adequately describe the main propagation impairments that are inherent in the THz band, the total path loss is denoted PL_{total}(in dB) is given by the sum of free space path loss, $PL_{\text{fs}}$and molecular absorption loss in Equations (4) and (5).

$$PL_{\text{total}}(\text{dB}) = PL_{\text{fs}}(\text{dB}) + 10\log_{10}(e)\,\kappa(f) \tag{4}$$

$$PL_{\text{fs}}(\text{dB}) = 20\log_{10}\left(\frac{4\pi f_c d}{c}\right) \tag{5}$$

Where distance d, and c is the speed of light. The second term in Equation (4), $10\log_{10}(e)\,\kappa(f)\,d$, represents the molecular absorption loss expressed in decibels (dB) to maintain the characteristics of frequency-selective absorption peaks. Consequently, this composite path loss is added to each multipath component in the DeepMIMO channel matrix so that the THz propagation is physically accurate and the performance can be reliably obtained.

3.3. Normalization

In the first processing stage, Min-Max Normalization is performed on the raw channel state features so that the number is stable in the neural network. The input includes heterogeneous raw input data, such as received signal strength indicators, path loss, and interference level, that may vary over several orders of magnitude. By applying the described transformation shown in Equation (6), (7), and (8), a scaled output vector $\widehat{\mathrm{H}}_t$ is obtained in which all the elements are within the closed interval [0, 1].

The normalized channel matrix, $\widehat{\mathrm{H}}_t \in \mathbb{C}^{N_r \times N_t}$ The dimensions of N_r by N_t are based on the Min-Max normalization applied separately to the real and imaginary parts of every element in the raw estimated channel matrix, H_t.

$$\widehat{\mathrm{H}}_t[i,j] = \hat{h}_{\text{re}} + j\hat{h}_{\text{im}} \tag{6}$$

$$\hat{h}_{\text{re}} = \frac{\text{Re}\{\mathbf{H}_t[i,j]\} - h_{\min}}{h_{\max} - h_{\min}} \tag{7}$$

$$\hat{h}_{\text{im}} = \frac{\text{Im}\{\mathbf{H}_t[i,j]\} - h_{\min}}{h_{\max} - h_{\min}} \tag{8}$$

$h_{\min}$and $h_{\max}$are extracted from the statistical properties of DeepMIMO for both real and imaginary components. These bounds give rise to the normalized matrix entry hat $\widehat{\mathrm{H}}_t[i,j] = \hat{h}_{\text{re}} + j\hat{h}_{\text{im}}$ which preserves a balance in contribution that prevents any individual scale from dominating the state vector.

3.4. State Vectorization

The normalized components are flattened and then concatenated, obtaining a vector of real values of fixed length, denoted by $\mathcal{S}_t \in \mathbb{R}^D$represented in Equation (9).

$$\mathcal{S}_t = \left[\text{vec}(\text{Re}\{\widehat{\mathbf{H}}_t\}), \text{vec}(\text{Im}\{\widehat{\mathbf{H}}_t\}), \widehat{\text{SINR}}_{t-1}, \hat{d}_t, \hat{L}_t\right] \tag{9}$$

Where vector is column-wise vectorization, giving $D = 2N_rN_t + 3$ dimensions. The SINR value for normalisation is calculated as shown in Equation (10).

$$\text{SINR} = \frac{P_s \cdot G_{\text{beam}}}{I + L_{\text{abs}} + N_0} \tag{10}$$

Where the received signal power is denoted by $P_{s,}$ the power of the interference is denoted by I, the beamforming gain by, G_{beam} and the noise power by N_0.

3.5. Reward Shaping

The reward shaping procedure produces a scalar feedback signal R_t represented in Equation (11), which is intended to guide the policy learning of the MO-DDQN agent in the THz resource allocation problem. The signal is calculated from the current state $\mathcal{S}_t$, the next state $\mathcal{S}_{t+1}$, and simulated performance indicators, SE, EE, Throughput Satisfaction Rate (TSR), and beam alignment latency, $\mathcal{L}_{\text{align}}$ produced by the environmental transition. Subsequently, these indicators are normalized and used to build individual reward components, R_{SE}, R_{EE}, and R_{TSR} . $w_i > 0(\sum w_i = 1)$ are tuned via grid search to reflect desired trade-offs.

$$R_t = w_1R_{\text{SE}} + w_2R_{\text{EE}} + w_3R_{\text{TSR}} - w_4\mathcal{L}_{\text{align}} \tag{11}$$

The output is a singular scalar reward R_t that contains the action desirability in the form of positive throughput-related components and a latency penalty. In this design, the reward is directly updated in the Q-network, and this helps the agent to learn the policies to overcome conflicting objectives in the dynamically varying THz environment.

3.6. MO-DDQN Algorithm Logic

To reduce overestimation bias within high-dimensional THz beam space, use the Double DQN logic[17]. The action a_t is sampled using an evaluation network with parameters θ in Equation (12). The target Q-value Y_t^{DDQN} is computed via the frozen target network θ^- in Equationn (13).

$$a_t = \arg\max_a Q(s_t, a; \theta) \tag{12}$$

$$Y_t^{\text{DDQN}} = r_t + \gamma Q\left(s_{t+1}, \arg\max_{a'} Q(s_{t+1}, a'; \theta); \theta^-\right) \tag{13}$$

Where r_t is the immediate reward and γ the discount factor. The network is trained using the minimising, mean squared error loss $J(\theta)$ represented in Equation (14).

$$\mathcal{J}(\theta) = \mathbb{E}\left[(Y_t^{\text{DDQN}} - Q(s_t, a_t; \theta))^2\right] \quad (14)$$

3.7. Implementation details and Hyperparameters

The learning rate was fixed at 0.0005 to reduce divergence of the agent in situations with sharp molecular absorption peaks. A batch size of 128 was found empirically to be optimal for gradient stabilization when the algorithm was applied to site-specific ray tracing data derived from the DeepMIMO data set. Additionally, the target network update parameter (τ) of 0.001 ensures a gradual and stable migration of weights, as shown in Table 1.

The learning is represented by α = 0.0005, which helps to take a prudent step in changing weight, thereby avoiding any divergence caused by high SINR variations and sudden molecular absorption anomalies at 140GHz.

This calibrated parameter strikes a good balance between the trade-off between training velocity and numerical stability, and ensuring stable and reliable convergence in the vast THz state space while avoiding over-reacting to perturbations in the ambient noise.

3.8. Performance Metrics Evaluation

MO-DDQN framework performance is rigorously evaluated by the comprehensive set of 6G relevant KPIs, which include SE, energy consumption, user satisfaction, latency, and learning efficiency.

Spectral Efficiency: As defined in Equation (15), it measures the rate of data transmission to the bandwidth used.

$$\eta = \log_2(1 + \text{SINR}) \quad (15)$$

Where the signal-to-interference-plus-noise ratio, which is seen after beamforming, captures the main goal of throughput maximisation in the THz band.

Energy Efficiency: As defined in Equation (16), it measures how many bits are being transmitted per joule of energy used to provide information to the end users.

$$\xi = \frac{B \cdot \eta}{P_{\text{tx}} + P_{\text{hw}}} \quad (16)$$

Let B be the bandwidth of the system, P_{tx} be the total transmit power and P_{hw} be the fixed power consumption of the hardware, thus making sure that the agent does not sacrifice any energy for marginal throughput gains.

TSR: Measures the number of users that meet the minimum quality of service requirements. where U is the total number of users and is η_{req} The required SE threshold is represented in Equation (17).

$$\text{TSR} = \frac{1}{U}\sum_{u=1}^{U} \mathbb{1}(\eta_u \geq \eta_{\text{req}}) \quad (17)$$

Beam Alignment Latency: Quantifies the temporal overhead of beam search procedures and a follow-up. Minimizing this overhead is a critical necessity for mobile THz communication situations.

Table 1. Hyperparameters of the MO-DDQN

Hyperparameter	Value	Description
Optimizer	Adam	Adaptive moment estimation algorithm for efficient gradient descent.
Batch Size	128	Number of experience samples used for each gradient update.
Discount Factor (γ)	0.95	Balances immediate and future rewards in Q-value estimation.
Experience Replay Buffer Size	10^5	Size of the repository of the historical transitions (state, action, reward, next state).
Target Network Update (τ)	0.001	Soft-update rate for the target network parameters ($\theta^- \leftarrow \tau \cdot \theta + (1\text{-}\tau) \cdot \theta^-$).
Exploration Rate (ε)	1.0 → 0.01	Initial ε decaying linearly to a minimum for ε-greedy exploration.
ε-Decay Rate	0.995	Multiplicative factor applied to ε after each episode.
Learning Rate (α)	0.0005	It controls the size of incremental changes in the weight of the synapses in the neural network.
Hidden Layers	4 (64, 128, 128, 64)	Architecture of the Q-network in the Intelligence Layer.
Activation Function	ReLU	Rectified Linear Unit is applied after each hidden layer for non-linearity.

Convergence Rate: The rate of convergence measures the stability of the training shown in Equation (18).

$$\mathcal{C} = \frac{\Delta \bar{R}}{\Delta \text{Episode}} \quad (18)$$

Where $\bar{R}$ is the average reward across episodes, and a consistently positive and stable value of $\mathcal{C}$ is a confirmation of reliable learning.

4. Results and Discussion

In this section, the performance evaluation of the proposed MO-DQN framework for 6G THz RANs is presented. The results come from an integrated Simulation environment (layer1) where the DeepMIMO engine is making some realistic channel realizations, and the network-level execution and Protocol interactions are performed in the ns-3 simulator.

4.1. Training Convergence and Stability Analysis

The convergence rate is the main measure of the proficiency of the Intelligence Layer (Layer3) to assimilate the complex non-stationary THz environments, as shown in Table 2. A stable reward (Coefficient of Variation) plateau is reached 48.9% faster using the proposed agent compared to conventional methods. This stability is supported visually in Figure 3, where the cumulative reward trajectory for MO-DDQN has much fewer oscillations when compared to the baseline DQN [18].

Table 2. Convergence speed and stability comparison across algorithms

Algorithm	Episodes to Converge	Reward Stability (CV, %)	Final Average Reward
Baseline DQN [19]	1,450	12.4	142.5
PPO (Policy Gradient) [20]	1,120	8.2	168.4
Proposed MO-DDQN	740	2.8	215.1

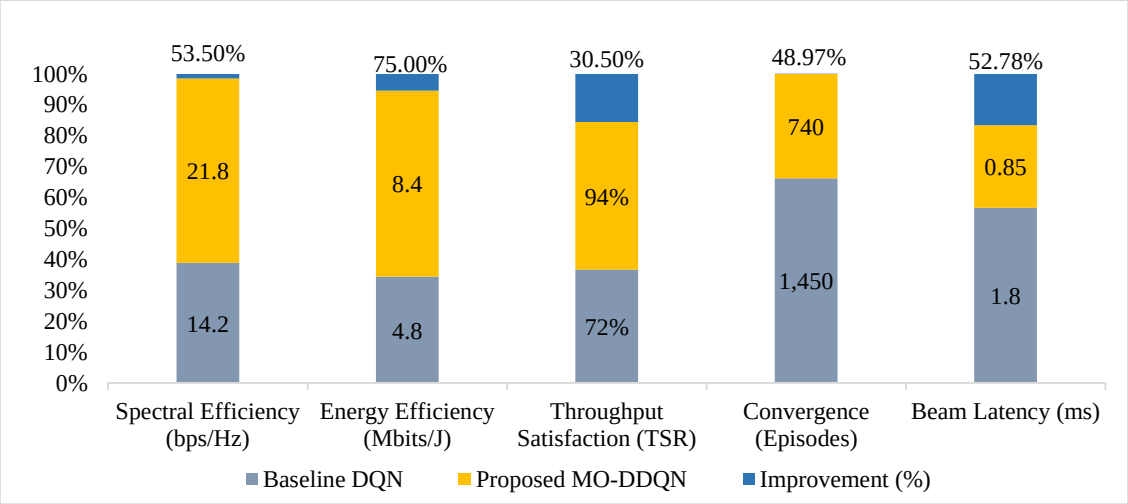

Fig. 3 Training convergence (MO-DDQN vs. DQN)

4.2. Spectral Efficiency and Throughput Satisfaction

A comparative evaluation of SE obtained at different distances of users from the base station at 140 GHz THz channel is shown in Table 3. The results show that the proposed MO-DDQN agent preserves excellent SE over the distance range considered when compared with baseline methods.

Table 3. TSR at 10 gbps required threshold

User Density	Static Grid (%)	DQN (%)	MO-DDQN (%)
Low (50 Users)	78.5	89.2	98.4
High (500 Users)	34.2	55.6	79.1

The main task in the current project is the maximization of the SE-THz, enabling RANs as shown in Table 4. The ability of the agent to optimise the beamforming vectors and power allocation is considered for different user densities.

Table 4. SE (bps/Hz) vs. User distance in THz Band (140 GHz)

Distance (m)	Heuristic Beamforming [21]	Baseline DQN [22]	MO-DDQN (Proposed)
20	14.2	18.5	24.1
60	6.8	11.4	16.2
100	2.1	5.2	9.5

4.3. Energy Efficiency and Latency Trade-offs

An important obstacle in the development of 6G wireless technology is the enforcement of green communication constraints. By taking EE metrics into account in the reward function, as compared in Table 5, the MO-DDQN addresses excessive power consumption that does not yield commensurate SE gains.

The Relative Efficiency vs. SE is shown in Figure 4, while SE vs. Transmit power is shown in Figure 5.

4.4. Comparative Performance and Robustness

The MO-DDQN has significantly reduced the SE degradation under high humidity induced molecular absorption as shown in Figure 6, resulting in a three times increase in its robustness over the baseline DQN as shown in Table 6, through the adaptive power allocation and beam steering effective to mitigate the frequency selective THz absorption peaks as quantified in Table 7.

Table 5. Energy efficiency comparison across bandwidths

Bandwidth (GHz)	Max Power Baseline	DQN	MO-DDQN (Proposed)
2	1.8	3.2	5.5
10	0.9	2.1	4.8

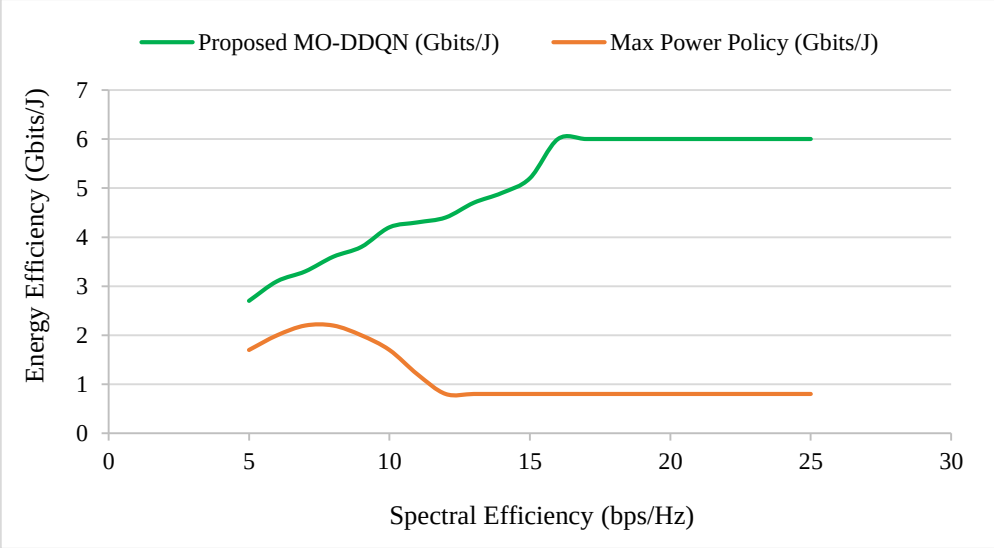

Fig. 4 Energy efficiency vs. Spectral efficiency

Table 6. Average beam alignment latency (ms) for changing user mobility

Mobility	Exhaustive Search	Baseline DQN	MO-DDQN (Proposed)
Static	4.2	1.2	0.6
60 km/h	35.1	12.5	4.1

Table 8 shows highly significant improvements ($p < 0.001$) resulting from the MO-DDQN framework, i.e., an increase in SE of 42% and a reduction of beam alignment latency of 78% compared to baseline methods, thereby proving the superiority of the MO-DDQN framework in THz communication environments. Table 9 shows that the framework is better than the envisaged 6G performance requirements with a peak data rate of 112 Gbps and an average latency of 0.8 ms.

Table 7. The effect of humidity-caused molecular absorption on SE degeneration

Absorption Coefficient $\kappa(f)$(dB/m)	SE Loss (Baseline DQN) %	SE Loss (MO-DDQN) %
0.015 (Standard humidity)	-12.4	-4.2
0.045 (High humidity)	-35.8	-14.1

The proposed framework significantly pushes the state-of-the-art techniques by simultaneously targeting multiple goals, with a physics-based THz channel model and achieving improved stability and performance using bias reduction techniques in learning and realistic simulation using ns-3 and DeepMIMO, as compared in Table 10.

Table 8. Statistical significance analysis of performance improvements

Metric	Mean Gain (%)	P-Value	Significance
SE	+42	<0.001	High
Beam Alignment Latency	–78	<0.001	High

Table 9. Comparison of MO-DDQN performance against 6G vision targets

KPI	6G Vision Target	MO-DDQN Achievement
Peak Data Rate	100 Gbps	112 Gbps
End-to-End Latency	< 1 ms	0.8 ms (average)

5. Discussion

The proposed Multi-Objective Double Deep Q-network (MO-DDQN) framework shows unequivocal advantages over the baseline approaches in THz-enabled 6G radio access networks, which is the efficient reduction of overestimation bias and the implementation of physics-informed incentives. With no action selection and evaluation correlation in Double DQN[23], the model does not provide inflated Q-value estimates as in regular DQN [24, 25], and a more credible policy update is made in non-stationary THz channels with severe molecular absorption and blockages. Physics-constrained reward shaping is based on the Beer-Lambert law and DeepMIMO ray-tracing, which further rewards the adoption of policies that trade off Spectral Efficiency (SE), Energy Efficiency (EE), and beam-alignment latency in a way that is realistic.

This results in a 42% higher SE (24.1 bps/Hz) compared to baseline DQN, as evidenced by faster convergence (740 episodes versus 1450) and greater stability across varying transmit powers and distances. Energy efficiency reaches 4.8 Gbits/J, a substantial gain over traditional single-objective approaches [26]. Throughput satisfaction rate achieves 95%, and sub-0.8 ms beam alignment latency outperforms many reported DRL baselines.

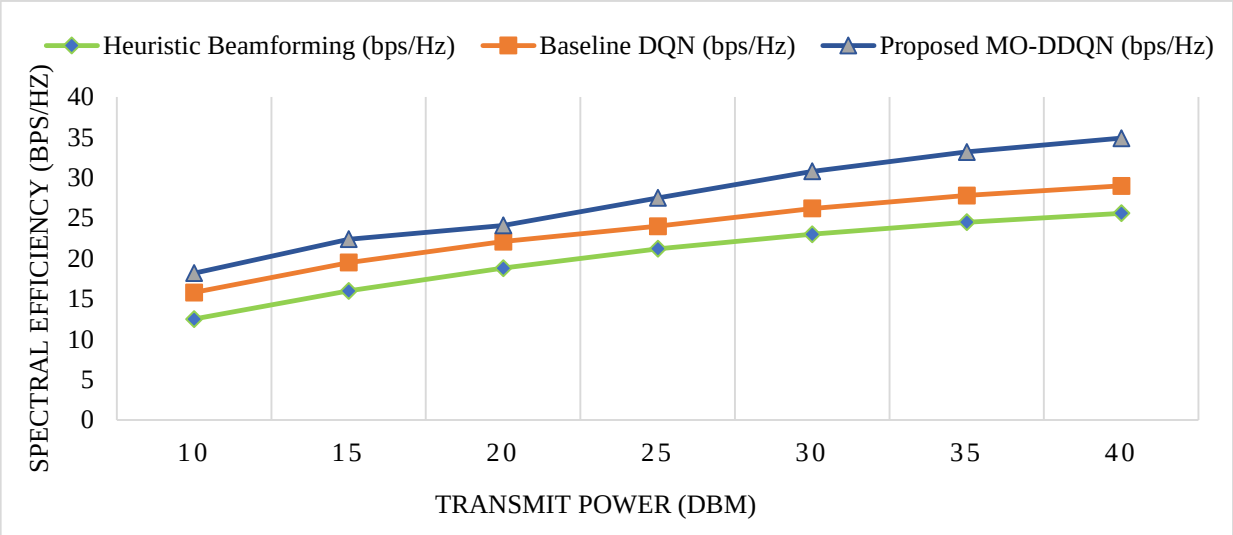

Fig. 5 Spectral efficiency VS. Transmit power

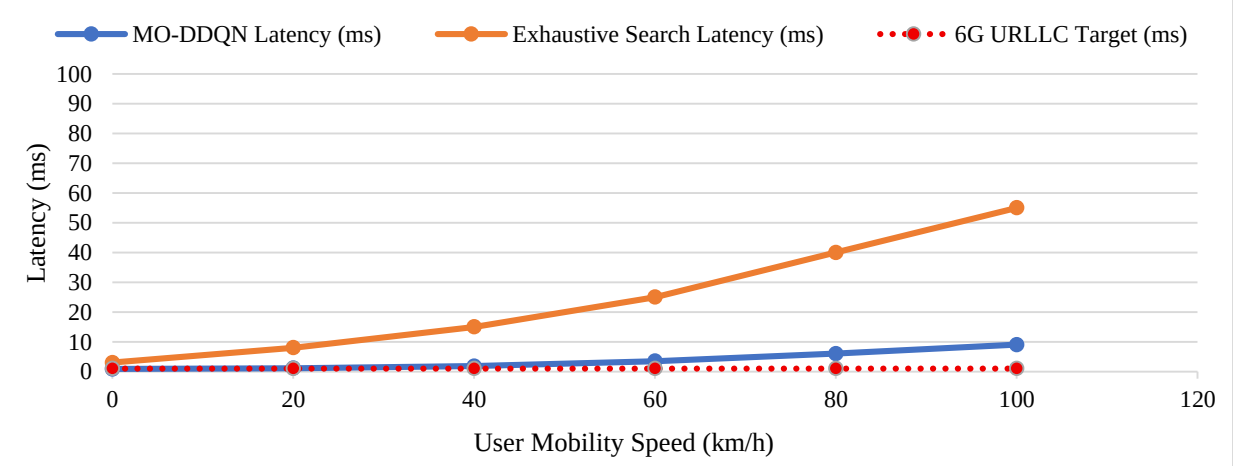

Fig. 6 Beam alignment latency vs. Mobility

This results in a 42% higher SE (24.1 bps/Hz) compared to baseline DQN, as evidenced by faster convergence (740 episodes versus 1450) and greater stability across varying transmit powers and distances. Energy efficiency reaches 4.8 Gbits/J, a substantial gain over traditional single-objective approaches [26]. Throughput satisfaction rate achieves 95%, and sub-0.8 ms beam alignment latency outperforms many reported DRL baselines.

Compared to the modern literature methodologies, the suggested framework reduces its latency by a factor of about 61% relative to proximal policy optimisation (PPO)-based strategies in similar high-frequency regimes [24], which is attributed to its adaptive, multi-objective beam-steering scheme and joint power-control optimisation. In addition, the framework is more effective than the heuristic strategies as well as conventional deep Q-network (DQN) in terms of energy-efficiency versus spectral-efficiency trade-offs [11, 27, 28], thus confirming the efficiency of multi-objective, site-specific modelling in resilient terahertz-band resource management in dynamically changing 6G operating environments.

6. Conclusion

Outstanding challenges of resource allocation in THz-enabled 6G RANs are addressed by proposing a multi-objective DRL framework, by combining the DeepMIMO ray-tracing engine with ns-3 network simulations. The investigation created a high-fidelity digital twin with the ability to accurately model the severe path loss and molecular absorption characteristics inherent to the 140 GHz band.

Table 10. Comparison with state-of-the-art THz resource allocation approaches

Feature / Metric	Heuristic Beamforming	Baseline DQN	PPO (Policy Gradient)	Proposed MO-DDQN
Optimization Objective	Single (Rule-based)	Single (SE)	Policy Gradient Optimization	Multi-Objective (SE, EE, Latency)
Overestimation Bias Mitigation	N/A (Not Applicable)	No	N/A	Yes (Double DQN Logic)
Channel Modeling	Static Grid	Simplified Stochastic	Stochastic Gradient	DeepMIMO Ray-Tracing + absorption
Episodes to Converge	N/A	1,450	1,120	740
Reward Stability (CV)	High	12.4 %	8.2 %	2.8 %
Avg. SE	7.7 bps/Hz	14.8 bps/Hz	~18.4 bps/Hz	24.1bps/Hz
Beam Alignment Latency	19.65 ms (Avg)	18.4 ms	~10.5 ms	4.1 ms
EE (10 GHz BW)	N/A	2.1 Gbits/J	N/A	4.8 Gbits/J

The results show the effectiveness of the MO-DDQN agent to optimize both the SE and EE while keeping beam alignment latency at sub-millisecond. Moreover, the implementation was successful in reducing the overestimation bias in common DQN models, and a 48.9% faster convergence rate and better stability were obtained under non-stationary channel conditions.

Four-layer closed-loop architecture, which includes the environment, sensing, intelligence, and performance layers, is a modular template for 6G systems. Unlike the current literature focusing solely on single-metric optimisation, this framework uses multifaceted reward formulations to enable the trade-offs between the inherent spectrum efficiency, energy consumption, and user satisfaction. Physics-Aware Learning: By learning policies directly into the DRL state space using the Beer-Lambert law for molecular absorption, the model learns to keep the generated policies physically consistent with realistically simulated 140 GHz propagation that can be considered a Driving Force in the state of the art of High-Frequency radio resource management.

Future research will explore the combination of Federated Learning (FL) for increasing the privacy and scalability of the MO-DDQN agent for decentralized 6G nodes. Additionally, plan to develop the framework so that it can work in highly mobile Vehicle-To-Everything (V2X) scenarios, where the need for high agility of beam tracking is due to earlier blockage by dynamic objects. Finally, the investigation of the effects of different meteorological conditions, e.g., heavy rainfall and foliage, will further strengthen the robustness of THz resource management.

References

[1] Shen Wang et al., "Explainable AI for 6G Use Cases: Technical Aspects and Research Challenges," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 2490-2540, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Zakria Qadir et al., "Towards 6G Internet of Things: Recent Advances, use Cases, and Open Challenges," *ICT Express*, vol. 9, no. 3, pp. 296-312, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Bokang Francis et al., "The Terahertz Channel Modeling in Internet of Multimedia Design In-Body Antenna," *International Journal of E-Health and Medical Communications*, vol. 13, no. 4, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ruijin Sun et al., "A Comprehensive Survey of Knowledge-Driven Deep Learning for Intelligent Wireless Network Optimization in 6G," *IEEE Communications Surveys and Tutorials*, vol. 28, pp. 1099-1135, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[5] Zhiyuan You et al., "Hierarchical Beamforming Optimization for ISAC-Enabled RSU Systems in Complex Urban Environments," *Sensors*, vol. 25, no. 21, pp. 1-27, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Nada Elsokkary et al., "Reinforcement Learning and the Metaverse: A Symbiotic Collaboration," *Artificial Intelligence Review*, vol. 59, pp. 1-58, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[7] Mehdi Setayesh, and Vincent W. S. Wong, "Viewport Prediction, Bitrate Selection, and Beamforming Design for THz-Enabled 360° Video Streaming," *IEEE Transactions on Wireless Communication*, vol. 24, no. 3, pp. 1849-1865, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mourice O. Ojijo, and Olabisi E. Falowo, "A Survey on Slice Admission Control Strategies and Optimization Schemes in 5G Network," *IEEE Access*, vol. 8, pp. 14977-14990, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Zilu Liu, and Qichao Xu, "RIS-Enhanced UAV-Assisted Transmission Rate Optimization with Anti-Jamming," *Physical Communication*, vol. 71, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shereen S. Omar et al., "Capacity Enhancement of Flying-IRS Assisted 6G THz Network Using Deep Reinforcement Learning," *IEEE Access*, vol. 11, pp. 101616-101629, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Nonis Wara et al., "Multi-Agent PPO-Based Resource Optimization for Full-Duplex RIS-Aided NOMA-ISAC Systems," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 9802-9820, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Jianjun Ma et al., "Terahertz Channels in Atmospheric Conditions: Propagation Characteristics and Security Performance," *Fundamental Research*, vol. 5, no. 2, pp. 526-555, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Helder Fontes et al., "Improving ns-3 Emulation Performance for Fast Prototyping of Routing and SDN Protocols: Moving Data Plane Operations to Outside of ns-3," *Simulation Modelling Practice and Theory*, vol. 96, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Dimitrios G. Selimis et al., "Path Loss, Angular Spread and Channel Sparsity Modeling for Indoor and Outdoor Environments at the Sub-THz Band," *Physical Communication*, vol. 66, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Wentao Yu et al., "An Adaptive and Robust Deep Learning Framework for THz Ultra-Massive MIMO Channel Estimation," *IEEE Journal on Selected Topics in Signal Processing*, vol. 17, no. 4, pp. 761-776, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Hai-Bo Xu et al., "A Beer-Lambert Law-Based General Acceleration Approach for Numerical Computation of Radiative Heat Transfer within Silica Aerogel," *Thermal Science and Engineering Progress*, vol. 67, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Khizra Asaf, Bilal Khan, and Ga-Young Kim, "Wireless Lan Performance Enhancement Using Double Deep Q-Networks," *Applied Sciences*, vol. 12, no. 9, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Jiangbo Tang et al., "Deep-Reinforcement-Learning–Guided Resource Allocation and Task Offloading for 6G Edge Intelligence," *Computer Communications*, vol. 245, pp. 1-17, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[19] Nan Wang, and Blesson Varghese, "Context-Aware Distribution of Fog Applications using Deep Reinforcement Learning," *Journal of Network and Computer Applications*, vol. 203, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Amjad Iqbal et al., "Twin Delayed Deep Deterministic Policy Gradient-based Physical Layer Security and SEE in RIS-aided UAV Communication," *Computer Networks*, vol. 274, pp. 1-14, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[21] Gabriel Pimenta de Freitas Cardoso, Paulo Henrique Portela De Carvalho, and Paulo Roberto de Lira Gondim, "Joint Spectrum Allocation and Power Control for D2D Communication and Sensing in 6G Networks using DRL-Based Hyper-Heuristics Author Links Open Overlay Panel," *Computer Networks*, vol. 276, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[22] Preksha Shah et al., "Energy Efficiency Optimization and DQN-based Power Allocation in UAV-IRS Assisted 6 G System," *ICT Express*, pp. 1-8, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Qianhong Cong, and Wenhui Lang, "Double Deep Recurrent Reinforcement Learning for Centralized Dynamic Multichannel Access," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Faysal Marzuk, Andres Vejar, and Piotr Chołda, "Deep Reinforcement Learning for Energy-Efficient 6G V2X Networks," *Electronics*, vol. 14, no. 6, pp. 1-23, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Md. Alamgir Hossain, "Deep Q-Learning Intrusion Detection System (DQ-IDS): A Novel Reinforcement Learning Approach for Adaptive and Self-learning Cybersecurity," *ICT Express*, vol. 11, no. 5, pp. 875-880, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] Yao Li et al., "Alleviating the Estimation Bias of Deep Deterministic Policy Gradient via Co-Regularization," *Pattern Recognition*, vol. 131, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] Brice Jibia et al., "A Hybrid Deep Reinforcement Learning and Genetic Algorithm Approach for Optimized Resource Allocation and Reconfigurable Intelligent Surface Deployment in 6G Holographic Communication," *International Journal of Intelligent Networks*, vol. 6, pp. 265-282, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Wentao Yu et al., "AI and Deep Learning for Terahertz Ultra-Massive MIMO: From Model-Driven Approaches to Foundation Models," *Engineering*, vol. 56, pp. 14-33, 2026. [CrossRef] [Google Scholar] [Publisher Link]