Wasit Journal of Engineering Sciences
Journal homepage: https://ejuow.uowasit.edu.iq
Vol. 13 No. 4 December 2025, pp.45-55
https://doi.org/10.31185/ejuow.v13i4.726

ISSN 2305-6932
Original Research

Double Deep Q-Network Techniques for Optimizing Performance of 6G Wireless Network

Yilmaz B. Kamal [1] , Ayad A. Abdulkafi[1]
[1]Department of Electrical Engineering, College of Engineering, Tikrit University, Tikrit, Iraq
Corresponding Author Email: yb230057en@st.tu.edu.iq

Received Jul.21, 2025 Revised Aug.29, 2025 Accepted Aug.31, 2025 Online Dec.1, 2025

ABSTRACT

This study introduces a Double Deep Q-Network (DDQN) optimization framework to improve massive MIMO-OFDM systems via reinforcement learning-driven adaptive parameter selection. It utilizes a dual network architecture to mitigate overestimation bias and incorporates dynamic optimization for power allocation, subcarrier fraction distribution, and modulation scheme selection across QAM-16, QAM-64, and QAM-128 configurations. Extensive simulations performed across Signal-to-Noise Ratio ranges from -5 to 35 dBm reveal substantial performance enhancements, with DDQN-augmented systems attaining 5-6 dB SNR savings for equivalent SE, a 50% increase in EE reaching 15.5-16 Gbps/W compared to conventional 10.5-11 Gbps/W implementations, and a 2.5 dB SNR reduction for a BER performance of 10^{-5}. The optimization framework ensures uniform parameter selection across diverse SNR conditions, facilitating a 40-50% increase in coverage through enhanced low-SNR performance while delivering a 5 dB SNR improvement in low-power operating scenarios. The study establishes a basis for intelligent communication systems that can autonomously adapt to 6G wireless networks, supporting ultra-reliable low-power communications and mobile edge computing applications.

Keywords: Sixth Generation Networks; Massive MIMO-OFDM; Double Deep Q-Network; SE.

1. Introduction

The swift advancement of 6G wireless communication systems necessitates optimization strategies to tackle the intricate challenges posed by massive MIMO-OFDM networks, where conventional systems exhibit considerable constraints in adaptive parameter selection and subpar performance under fluctuating channel conditions. 6G networks require high efficiency, reliability, and adaptability for wireless communication systems. Due to 6G networks' dense, heterogeneous, and dynamic nature, conventional optimization methods struggle to meet performance objectives, necessitating intelligent and adaptive methods for real-time decision-making [1]. Integrating the Orthogonal Frequency Division Multiplexing (OFDM) is a pivotal waveform technique owing to its robustness against frequency-selective fading and effective spectrum use [2]. Traditional OFDM systems employ static parameter setups that ineffectively adapt to the frequently hesitant channel circumstances typical of 6G deployments. This constraint is especially evident when accommodating various modulation schemes under different signal-to-noise ratio (SNR) situations [3]. Progression required essential advances in physical layer technology to meet the extraordinary performance demands of 6G. OFDM waveforms, although effective in prior generations, must undergo substantial evolution to satisfy these requirements [4]. MIMO systems use multiple antennas for transmission and reception, improving wireless communication quality and capacity. Spatial diversity enhances reliability, achieving significant capacity increases. FDMA systems use techniques like MIMO, error control, and interference. Massive MIMO improves spectral and EE by using many antenna elements [5][6].

Machine Learning (ML), particularly Reinforcement Learning (RL), offers solutions to such problems through agent training and adaptation facilitated by environmental discovery and communication systems [7]. RL networks effectively tackle intricate decision-making problems defined by extensive state spaces in deep Q networks (DQNs), and the Double Deep Q-Network (DDQN) variation improves performance by addressing the overestimation bias present in conventional DQN implementations [8].

The selection of massive MIMO-OFDM parameters is performed using a Markov Decision Process (MDP), based on the observed channel quality and performance indicators, the intelligent DDQN agent gains experience and can make sequential judgments regarding subcarrier distribution and power allocation [9]. Resource allocation and adaptive waveform design are essential for optimizing SE and ensuring dependability under varying operational conditions [10].

6G wireless networks face challenges in achieving efficiency, reliability, and adaptability in complex communication environments. Conventional mMIMO-OFDM systems have degraded performance due to static transmission parameters. Existing wireless optimization frameworks focus on isolated performance metrics without considering trade-offs. RL techniques, such as the DDQN, are underexplored for optimizing power allocation and subcarrier selection across different modulation schemes. This study presents a framework for dynamic wireless communication optimization using DDQN-based RL, deep neural network (DNN) and Markov Dissection Processor. It enhances antenna size in mMIMO-OFDM systems, facilitates seamless power and subcarrier allocation, and maintains high-performance robustness across various SNR ranges. This research proposes a DDQN framework to optimize mMIMO-OFDM systems in 6G wireless communications, focusing on adaptive power allocation and subcarrier distribution based on channel conditions to enhance Spectral Efficiency (SE), Energy Efficiency (EE), and Bit Error Rate (BER).

This paper is organized as follows: Section 2 details our proposed approach, including the system model, DDQN formulation, and implementation architecture. Section 3 describes the results and critical discussion. Section 4 concludes the paper and suggests directions for future research.

2. System Model:

We examine DDQN as the foundation for a 6G wireless communication system that enhances performance in single-cell massive MIMO-OFDM configurations, integrating it with the channel model of Adaptive White Gaussian Noise (AWGN), where the base station is designated as the agent, and additional assumptions are detailed in the main table. 1. DDQN facilitates the allocation of subcarriers and power by networks, hence improving performance metrics including spectrum efficiency, EE, and BER.

2.1 Research methodology:

This study examines the characteristics of massive MIMO-OFDM and DDQN, as well as the agent's work strategy, as shown in Figure 1. It integrates DDQN with AWGN wireless network massive MIMO-OFDM to evaluate enhancements in SE, EE, and BER. It transforms random states into non-linear relationships and trains a DNN to execute actions, enhancing performance through a reward approach.

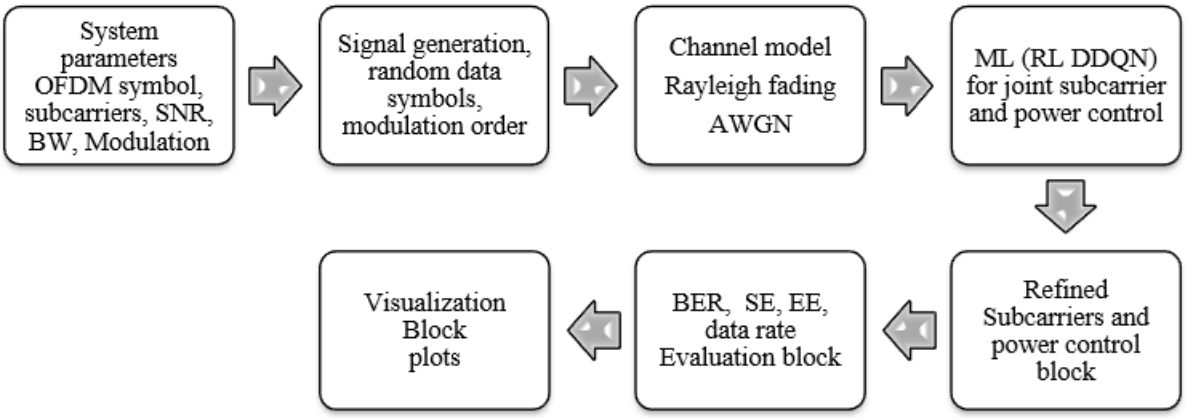

Figure 1. Flowchart of DDQN-Based massive MIMO-OFDM Wireless Networks.

The proposed system uses dynamic subcarrier allocation and adaptive massive MIMO-OFDM modulation. The transmitted signal can be written as follows [11]:

$$x(t) = \sum_{k=0}^{N-1} \sum_{n=-\infty}^{\infty} s_{k,n} g(t - nT) e^{j2\pi f_k (t-nT)} \quad (1)$$

where $s_{k,n}$ is the complex symbol sent on the k-th subcarrier during the n-th symbol period, $g(t)$ is the pulse shaping filter, T the symbol duration, and $f_k = f_0 + k\Delta f$ is the subcarrier frequency. The signal received post-synchronization and cyclic prefix removal is expressed as:

$$\mathrm{y(t)} = \sum_{\mathrm{k=0}}^{\mathrm{N-1}} \mathrm{H_k s_k e^{j2\pi f_k t}} + \mathrm{n(t)} \quad (2)$$

Here, H_k signifies the frequency domain channel coefficient for the k-th subcarrier, whereas $n(t)$ represents additive white Gaussian noise with variance σ^2. The system employs adaptive modulation on subcarriers according to real-time channel conditions, with modulation order M_k [12]. The massive MIMO signals can be mathematically represented as shown in Eq. (3), which computes the received signal Rr and H delineates the channel functions, comprising (number of receiving antennas (Nr) × number of transmitting antennas (Nt) elements, with (n) symbolizing the noise [13].

$$\mathrm{Rr = HX + n,} \quad (3)$$

The formulas for matrices will be written as follows:

- **Spectral Efficiency (η_{SE}):** SE, measured in (bits/s/Hz), refers to the information rate within a specific bandwidth in a communication system. Higher-order modulation and a low code rate can improve this [14].

$$\eta_{SE} = \log_2(1 - \mathrm{N_r N_t} \gamma_k) \quad (4)$$

- **Energy Efficiency (EE):** in a wireless communication system, it is considered as the ratio of power used to the number of precisely sent bits. According to the definition above, and depending on many power consumption terms like a "constant power" (Pc), which refers to a consistent quantity that incorporates both the energy used for control signals and the energy used by baseband processors and backhaul equipment for their load-independent operations [15].

$$P_c = P_0 + P_{RF} + \frac{P_{PA}}{\eta_{PA}} + P_{SP} \quad (5)$$

$$EE = \frac{B.\eta_{SE}}{P_C} \quad \text{EE units are (bits/J)} \quad (6)$$

- **Bit Error Rate (BER_k):**

The system works correctly for a single-carrier environment, implementing modulation order M_k. The M-ary QAM for the k-th subcarriers is [16]:

$$\mathrm{BER_k} = \frac{\sqrt{\mathrm{M}}-1}{\sqrt{\mathrm{M}}\log_2 \sqrt{\mathrm{M}}} \mathrm{erfc}\left(\sqrt{\frac{3 \cdot \log_2 \mathrm{M} \cdot \gamma_k}{2(\mathrm{M}-1)}}\right) \quad (7)$$

2.2 DDQN-based MIMO-OFDM system:

Figure 2 represents the systematic view of communication and processing in RL DDQN-based systems. Initially, the random state is at the MIMO-OFDM system, which train the DNN based on the corresponding DDQN algorithm. The DDQN agent functions in a closed-loop manner, persistently monitoring the state space (SNR, current BER, modulation parameters), choosing actions (power allocation, subcarrier assignment), and obtaining incentives based on multi-objective performance metrics. The Markova Decision Process (MDP) is based on DDQN and relied on several components: (S, A, π, R), S represents the state space, which includes all possible states and refers to the environment in which the agent operates. A represents the action space, which refers to all possible decisions the agent can make from this space. If an agent observes a state s_t and takes a decision a_t for a time instance t, the transition probability $P_a(s_t, s_{t+1})$ is likelihood that the current state s_t becomes s_{t+1} for the subsequence step t. π refers to the agent's decision-making strategy. The likelihood for taking action a_t on the state s_t by an agent is given as $\pi(s_t, a_t)$, where $a_t \in A, s_t \in S$, and $\sum_{a_t \in A} \pi(s_t, a_t) = 1, \forall s_t \in S$. Furthermore, R represents the amount of benefit an agent receives based on the current state and action. When an agent takes any action, there will be a reward, and this will determine which action the agent will take to maximize the goal value. Suppose an agent performs an action a_t at state s_t; the resulting reward can be expressed as r_{t+1}.

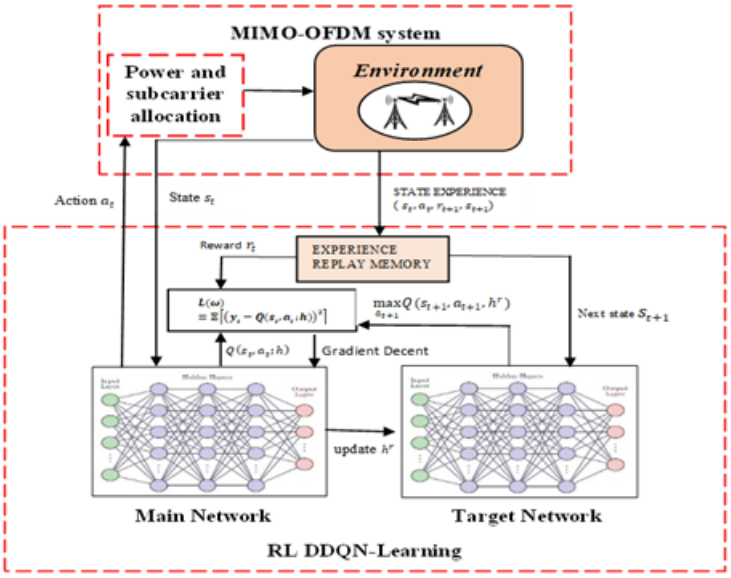

Figure 2. Communication and Processing in DDQN-Based massive MIMO-OFDM system.

In the proposed single-agent deep reinforcement learning algorithm, we delineate the agent, states, actions, and reward function as follows.

- **Agent**: MIMO-OFDM Base station.
- **The state:** The state space is constructed to capture relevant system and channel characteristics while maintaining dimensionality suitable for practical implementation:

$$S = \{[\gamma_{dB}, BER_{current}, M] \mid \gamma_{dB} \in \{0,35\}, BER_{current} \in \{0,1\}, M \in \{16,64,128\}\} \quad (8)$$

Where $(\gamma_{dB}, BER_{current}, M)$ represent SNR, current BER, and modulation order. This approach produces a three-dimensional continuous state space corresponding to the system's operating condition.

- **The action**: Consists of a discrete set of configuration parameters:

$$A = \{(P_k, \alpha_{SC}) \mid P_k \in \mathcal{P}, \alpha_{SC} \in \mathcal{F}\} \quad (9)$$

Where $\mathcal{P}$ represents the set of possible power scaling factors $P_k \in \{0.2, 0.4, \ldots., 1\}$and $\mathcal{F}$ denotes the set of available subcarrier fractions $\alpha_{SC} \in \{0.2, 0.4, \ldots., 1\}$. This results in a discrete action space with $|A| = |\mathcal{P}| \times |\mathcal{F}|$ distinct actions, each representing a unique configuration of power allocation and subcarrier utilization.

- **The reward:**

$$R_t = w_1 \cdot f(\eta_{SE}) + w_2 \cdot f(EE) - w_3 \cdot f(BER_k) \quad (10)$$

Where:

- w_1, w_2, w_3 are weighting factors that determine the relative importance of each objective
- $f(\cdot)$ represents normalization for each metric

The system implements adaptive subcarrier allocation and power scaling factor adjusts transmit power across all subcarriers, are utilized based on channel conditions. For $\alpha_{SC} < 1$, the active subcarriers $N_{active} = \lfloor \alpha_{SC} \cdot N \rfloor$ are selected according to Eq. (12) [17]:

$$\mathcal{K}_{active} = \{k_1, k_2, \ldots, k_{N_{active}}\} \subset \{0,1, \ldots, N-1\} \quad (11)$$

Additionally, resulting in an effective SNR:

$$\gamma_{eff,k} = \gamma_k \cdot P_k \quad (12)$$

The system operates in a three-dimensional state space with 25 action states for steady learning convergence. The reward function uses a balanced optimization approach, prioritizing SE, EE, and BER. The system allocates subcarriers based on channel gain and power scaling, balancing trade-offs between known configurations and potential superior alternatives. This technique ensures EE and SE efficient system operation, maintains communication quality, and adheres to power limitations [18].

The EE optimization can be written as:

$$\left.\begin{aligned} &\max_{P_k,\alpha_{SC}} EE(P_k, \alpha_{SC}), \\ &\text{subject to } BER(P_k, \alpha_{SC}) \le BER_{target} \\ &0.2 \le P_k \le 1.0 \\ &0.2 \le \alpha_{SC} \le 1.0 \\ &\sum_{i=1}^{N_t} \sum_{k \in \mathcal{K}_{active}} P_{i,k} \le P_{total} \end{aligned}\right\} \quad (13)$$

The SE optimization can be expressed as:

$$\left.\begin{aligned} &\max_{P_k,\alpha_{SC}} \eta_{SE}(P_k, \alpha_{SC}) \\ &\text{subject to } BER(P_k, \alpha_{SC}) \le BER_{target} \\ &0.2 \le P_k \le 1.0 \\ &0.2 \le \alpha_{SC} \le 1.0 \\ &\sum_{i=1}^{N_t} \sum_{k \in \mathcal{K}_{active}} P_{i,k} \le P_{total} \end{aligned}\right\} \quad (14)$$

The SE of DDQN mMIMO-OFDM systems can written as:

$$\eta_{SE_{DDQN}} = \log_2\left(1 - N_r N_t \gamma_{k\_DDQN} \cdot \alpha_{SC}\right) \quad (15)$$

Also, the power and EE can be express as follow:

$$P_{c\_DDQN} = P_0 + N_t . P_{RF} + \frac{P_{PA}}{\eta_{PA}} . N_t . P_k + P_{SP} \quad (16)$$

$$EE_{DDQN} = \frac{B.\eta_{SE}}{P_{c\_DDQN}} \quad (17)$$

BER can be express as following:

$$BER_{DDQN} = \frac{\sqrt{M}-1}{\sqrt{M}\log_2 \sqrt{M}} \operatorname{erfc}\left(\sqrt{\frac{3\cdot\log_2 M \cdot \gamma_{eff,k}}{2(M-1)}}\right) \quad (18)$$

The limitations of SE, EE, and BER of the DDQN mMIMO-OFDM systems can be articulated as:

$$\left.\begin{aligned} &\eta_{SE_{DDQN}}(\beta_P, \alpha_{SC}) \ge \eta_{SE_{min}} \\ &EE_{DDQN}(\beta_P, \alpha_{SC}) \ge EE_{min} \\ &BER_{DDQN} < BER_{k_{min}} \end{aligned}\right\} \quad (19)$$

Constraints require that the power and subcarrier allocation to BS be nonnegative and that the overall power and subcarrier assigned by the BS does not exceed its available capacity.

A deep Q-network method is used to determine maximum policies using DNNs instead of Q-value tables in deep learning [54]. The environment state s_t is input to DNN to execute the predicted. Q-values in the form of $Q(s_t, a_t; h)$, $a_t \in A$ where h refer the weights of the DNN. Moreover, Agent acquires the optimal strategy for Q-function optimization, which entails minimizing the loss function L, as defined below [19]:

$$L(\omega) = \mathbb{E}\left[\left(y_t - Q(s_t, a_t; h)\right)^2\right] \quad (20)$$

In the training phase, the DNN parameters are sequentially updated for Q-function. Optimization. This repeated update process of DNN parameters. h is expressed as eq. below [20]:

$$h = h + \alpha\mathbb{E}\left[\left(y_t - Q(s_t, a_t; h)\right)\nabla Q(s_t, a_t; h)\right] \quad (21)$$

The agent retains experience data in an experience replay pool as a tuple ($s_t, a_t, r_{t+1}, s_{t+1}$) instead of utilizing a singular experience data point for training throughout each iteration. Consequently, the training of the DNN entails the random selection of a mini-batch of samples from the experience replay memory [21]. Where the DQN algorithm utilizes two Q networks: online and target networks, both online and target networks exhibit analogous designs. However, they carry distinct weights. The weights of the online network are denoted as h^m, whereas the weights of the target network are designated as h^r. The goal Q-value presented as follows [22]:

$$y_t = r_{t+1} + \lambda \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}, h^r) \quad (22)$$

The DDQN algorithm for massive MIMO-OFDM utilizes two maximum function estimators to differentiate between action selection and value evaluation, thereby replicating overestimation to achieve a balanced Q-value. The estimation weights are utilized to determine the optimal action in this process. The greedy policy strategy value is determined by the current value represented by, whereas the second set of weights, as described in (23) [23]:

$$y_t = r_{t+1} + \lambda Q\left(s_{t+1}, \arg\ \underset{a_{t+1}}{max} Q(s_{t+1}, a_{t+1}; h^m); h^r\right) \quad (23)$$

The progression of DDQN in wireless communications encounters obstacles in reward design, optimization of DNN architecture, and sample efficiency, with insufficient study about its efficacy in massive MIMO-OFDM systems for 6G applications.

3. Simulation Results:

The core simulation components include a massive MIMO-OFDM signal processing chain, a DDQN agent implementation with experience replay and target network, and a custom environment interface for RL training and evaluation for calculating enhanced SE, EE, and BER. Key system parameters are configured in table 1.

Table.1 Main System Parameters.

Parameter	Value
Discount Factor	0.97
Learning Rate	10^{-3}
Batch Size	128
Experience Buffer Size	$5*10^5$
Epsilon Decay	0.998
Number of Subcarriers (N)	1024
Cyclic Prefix Length	N/4
System Bandwidth	6.4 GHz
OFDM Symbols	512
Transmit Power	p = 5W [24]
Fixed Power	P_0 = 0.1W [8]
Power per RF Chain	P_{RF} = 0.1W [25]
Power Amplifier per Antenna	P_{PA} = 0.7W
Power Amplifier Efficiency	η_{PA} = 0.45; ($0 < \eta_{PA} \leq 1$), (45%) [15]
Signal Processing Power	P_{SP}= 0.5W [15]

These parameters were selected to reflect the realistic capabilities of a 6G system while ensuring computational feasibility for extensive simulations. Three modulation techniques (16-QAM, 64-QAM, and 128-QAM) are used in the 256×256 MIMO-OFDM system with 512 subcarriers. With 0.5 dB resolution, performance is assessed over SNR ranges from -5 to 35 dBm. Trained across 800 episodes, the DDQN agent chooses the best combinations of subcarrier allocation (0.2-1) and power scaling (0.2-1). Simulation results are categorized into three segments. First, we analyze spectrum efficiency; second, we examine EE; and lastly, we review BER performance, as follows:

4.1. Spectral Efficiency:

Figure 4. Illustrates DDQN optimization framework has significantly improved SE performance in 6G wireless communications. The DDQN-enhanced schemes consistently outperform conventional schemes across all modulation orders, demonstrating the effectiveness of machine learning-based optimization in adaptive

wireless systems. The DDQN framework offers significant benefits at moderate to high SNR levels, where adaptive power allocation and subcarrier distribution features provide substantial benefits. The DDQN optimization results in uniform SNR reductions of 5-6 dB across all modulation schemes, enabling 64-QAM to achieve SE levels that often require 16-QAM at higher power levels. This leads to an improved energy economy and increased coverage capabilities for mobile edge computing and IoT applications. Higher-order modulations (128-QAM) offer significant enhancements, with DDQN optimization enabling practical application at SNR levels, accommodating varied quality-of-service demands in 6G networks.

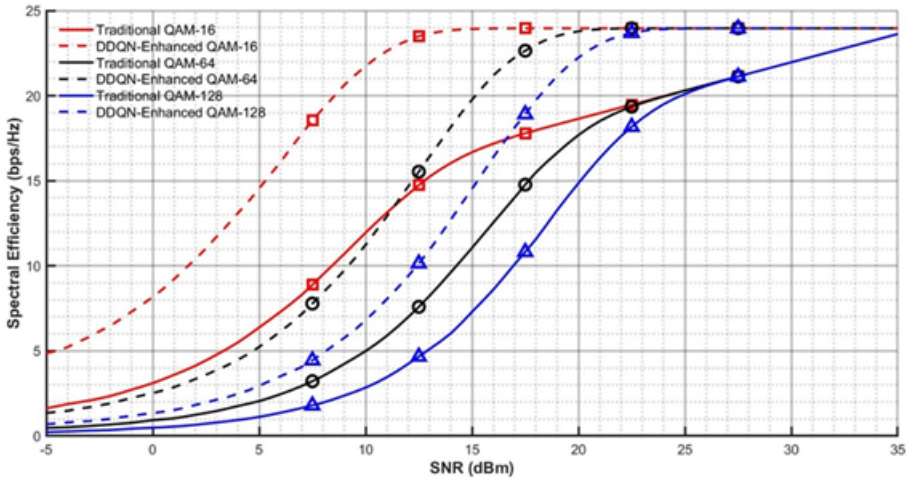

Figure 4. SE Comparison Between Traditional and DDQN Algorithm.

The following table 2. Summarizes the performance points extracted from the SE analysis:

Table 2. SNR Gain at Every Modulation Scheme (bits/Hz).

Modulation Scheme	SNR (dB)	Traditional SE (bps/Hz)	DDQN SE (bps/Hz)	SNR Gain (dB)	Performance Improvement
16-QAM	10	~9.0	~18.5	~5-6	105%
16-QAM	15	~17.5	~23.5	~5	34%
64-QAM	15	~7.5	~15.5	~6	107%
64-QAM	20	~15.0	~22.5	~5	50%
128-QAM	20	~5.0	~11.0	~6	120%
128-QAM	25	~11.0	~19.0	~5	73%

4.2. Energy Efficiency Analysis:

The study assesses EE in conventional and DDQN-enhanced QAM modulation techniques in massive MIMO-OFDM systems, providing insights into the optimization potential of machine learning-augmented wireless communication systems. EE curves show trade-offs between modulation complexity and energy performance, challenging traditional beliefs about higher-order modulation's benefits. 16-QAM schemes achieve exceptional EE maxima at 15-16 Gbps/W, surpassing 64-QAM and 128-QAM alternatives under most operating conditions. DDQN enhancement yields significant improvements across all modulation orders, with the most critical gains in the 10-25 dBm SNR band. However, the effectiveness varies with modulation complexity, suggesting lower-order modulations are more suitable for DDQN optimization. As shown in Figure 5.

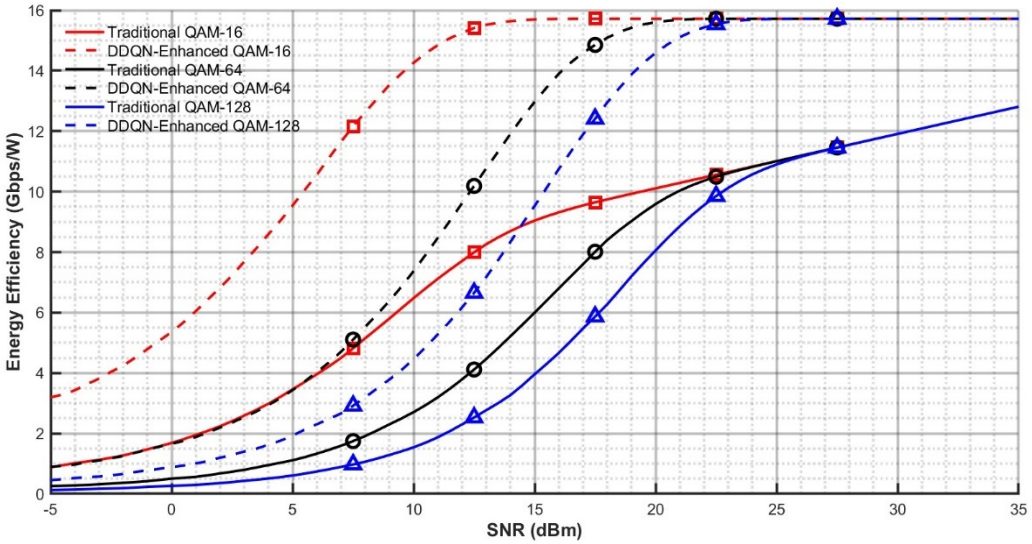

Figure 5. EE Comparison Between Traditional and DDQN Algorithm.

The analysis shows that DDQN optimization is effective within the 5-15 dBm SNR range in wireless systems. It enables lower-order modulations to achieve EE levels, increases network coverage, and reduces power consumption. The advantages are most noticeable in difficult propagation situations, making it ideal for 6G deployments focusing on widespread connectivity and EE. The following table 3. Summarizes the critical performance points extracted from the EE analysis:

Table 3. EE (Gbps/W) at Different SNR

SNR (dB)	Traditional QAM-16	DDQN QAM-16	Traditional QAM-64	DDQN QAM-64	Traditional QAM-128	DDQN QAM-128
-5	~0.2	~0.5	~0.1	~0.3	~0.05	~0.2
5	~2.0	~5.0	~1.5	~3.0	~0.8	~2.5
10	~5.0	~12.0	~2.0	~5.0	~1.0	~3.0
15	~8.0	~15.0	~4.0	~10.0	~2.5	~7.0
35	~10.5	~15.5	~12.0	~15.5	~12.5	~15.5

4.3. BER Performance

BER performance curves achieve significant gains in massive MIMO-OFDM systems with DDQN optimization under all QAM modulation orders (16-QAM, 64-QAM, and 128-QAM). The research demonstrates that DDQN-based systems outperform traditional designs consistently to provide approximately 2.5-5 dB SNR gains over the operating range of -5 to 35 dBm. Conventional 16-QAM needs approximately 20 dBm to achieve a BER of 10^{-3}, whereas the DDQN-enhanced counterpart achieves the same at 15 dBm, an improvement of 5 dB The relative improvements for higher-order modulations are impressive, with 128-QAM reaping huge gains through DDQN optimization, with the possibility of practical application at relatively modest SNR conditions when conventional methods cannot ensure tolerable error rates. Significant performance improvement is observed in the 10-20 dBm range of SNR, which is by normal wireless operating conditions, and hence, the DDQN architecture is suitable for practical applications. In high SNR regimes (25+ dBm), all the boosted methods converge to 10^{-5} or lower BER values, and the DDQN-boosted 256-QAM reaches a BER of 10^{-5} at 29.5 dBm, whereas in standard implementations it is at 32 dBm. The consistent performance gain for all modulation orders and SNR conditions confirms the validity of the deep reinforcement learning solution for power allocation adaptation and parameter tuning in 6G wireless communication systems for ultra-reliable low-power communications with resilient performance under adverse propagation environments. As illustrates in the Figure 6.

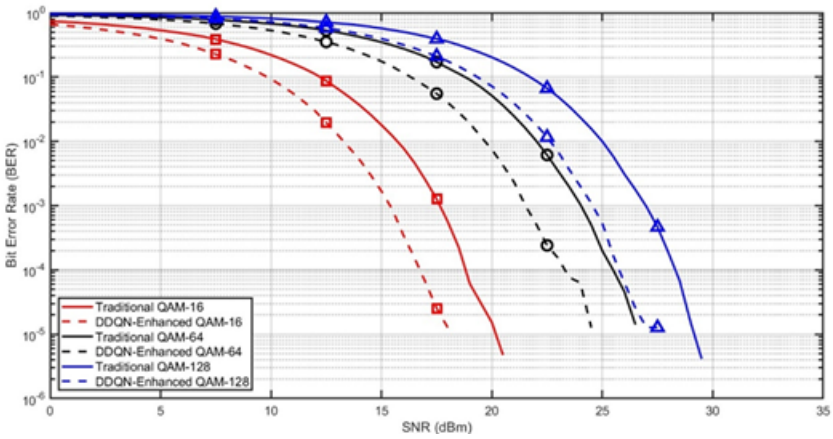

Figure 6. BER Performance Comparison Between Traditional and DDQN Algorithm.

The following table 4 summarizes the critical performance points extracted from the BER analysis:

Table 4. BER Gain at Different SNR (dB).

SNR (dB)	Traditional 16-QAM	DDQN 16-QAM	Traditiona l 64-QAM	DDQN 64-QAM	Traditional 128-QAM	DDQN 128-QAM
5	$\sim 10^{-1}$	$\sim 5\times10^{-2}$	$\sim 3\times10^{-1}$	$\sim 10^{-1}$	$\sim 5\times10^{-1}$	$\sim 2\times10^{-1}$
15	$\sim 10^{-2}$	$\sim 10^{-4}$	$\sim 2\times10^{-2}$	$\sim 10^{-3}$	$\sim 8\times10^{-2}$	$\sim 10^{-2}$
25	$\sim 10^{-4}$	$\sim 10^{-5}$	$\sim 10^{-3}$	$\sim 10^{-5}$	$\sim 5\times10^{-3}$	$\sim 10^{-4}$
30	$\sim 10^{-5}$	$\sim 10^{-5}$	$\sim 10^{-4}$	$\sim 10^{-5}$	$\sim 10^{-3}$	$\sim 10^{-5}$

4. Conclusion

DDQN optimization framework to improve mMIMO-OFDM systems via RL-driven adaptive parameter selection. The integration of MDP formulation with deep neural network topologies has transformed adaptive parameter optimization in dynamic wireless environments, reducing overestimation bias and guaranteeing robust convergence across varying SNR conditions. Findings show significant performance enhancements, with DDQN-augmented systems achieving 5-6 dB SNR reductions for equivalent SE and a 50% increase in EE, reaching 15.5-16 Gbps/W relative to traditional implementations. The BER analysis reveals a requirement for a 2.5 dB reduction in SNR for advanced systems. The future research should develop multi-agent DDQN frameworks for distributed massive MIMO systems, enabling cooperative learning for intelligent interference management and resource allocation.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Funding Information

No funding was received from any financial organization to conduct this research

Author Contributions

Yilmaz B. Kamal contributed to the conceptual design of the study, implementation of the reinforcement learning algorithm, simulation development, and preparation of the original manuscript draft.Ayad A. Abdulkafi provided supervision, guidance on methodological structure, critical manuscript revisions, and contributed to the final editing and review process. Both authors read and approved the final manuscript.

Acknowledgments

The authors express their gratitude to Wasit University/ College of Engineering/Electrical Engineering department in Alkut-Wasit-Iraq for supporting this study. In addition, many thanks to Ayad A. Abdulkafi for their advice for the academic writing of this paper.

References

[1] L. Li, "Research on future 6G green wireless networks," *Green Technologies and Sustainability*, vol. 3, no. 2, p. 100156, 2025, doi: 10.1016/j.grets.2024.100156.

[2] X. You *et al.*, "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," Jan. 01, 2021, *Science in China Press*. doi: 10.1007/s11432-020-2955-6.

[3] Z. Hu *et al.*, "Highly Efficient MIMO-OFDM with Index Modulation," *Wirel Commun Mob Comput*, vol. 2022, 2022, doi: 10.1155/2022/7399457.

[4] J. Hou, J. M. Purushothama, H. Fan, C. Song, Y. Ding, and M. Sellathurai, "Energy efficient time-modulated OFDM directional modulation transmitters," *Microw Opt Technol Lett*, vol. 65, no. 1, pp. 5–13, 2023, doi: 10.1002/mop.33438.

[5] H. Kasban, S. Hashima, S. Nassar, E. M. Mohamed, and M. A. M. El-Bendary, "Performance enhancing of MIMO-OFDM system utilizing different interleaving techniques with rate-less fountain raptor code," *IET Communications*, vol. 16, no. 20, pp. 2479–2491, Dec. 2022, doi: 10.1049/cmu2.12503.

[6] I. H. Ahmed and A. A. Abdulkafi, "Energy-Efficient Massive MIMO Network," *Tikrit Journal of Engineering Sciences*, vol. 30, no. 3, pp. 1–8, 2023, doi: 10.25130/tjes.30.3.1.

[7] A. Zaki, A. Métwalli, M. H. Aly, and W. K. Badawi, "Wireless Communication Channel Scenarios: Machine-Learning-Based Identification and Performance Enhancement," *Electronics (Switzerland)*, vol. 11, no. 19, 2022, doi: 10.3390/electronics11193253.

[8] E. J. Han, M. Sengly, and J. R. Lee, "Balancing Fairness and Energy Efficiency in SWIPT-Based D2D Networks: Deep Reinforcement Learning Based Approach," *IEEE Access*, vol. 10, pp. 64495–64503, 2022, doi: 10.1109/ACCESS.2022.3182686.

[9] F. H. Juwono and R. Reine, "Future OFDM-based Communication Systems Towards 6G and Beyond: Machine Learning Approaches," *Green Intelligent Systems and Applications*, vol. 1, no. 1, pp. 19–25, 2021, doi: 10.53623/gisa.v1i1.34.

[10] Z. Liwen, F. Qamar, M. Liaqat, M. Nour Hindia, and K. Akram Zainol Ariffin, "Toward Efficient 6G IoT Networks: A Perspective on Resource Optimization Strategies, Challenges, and Future Directions," *IEEE Access*, vol. 12, no. April, pp. 76606–76633, 2024, doi: 10.1109/ACCESS.2024.3405487.

[11] J. M. Hamamreh, A. Hajar, and M. Abewa, "Orthogonal frequency division multiplexing with subcarrier power modulation for doubling the spectral efficiency of 6G and beyond networks," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 4, pp. 1–31, 2020, doi: 10.1002/ett.3921.

[12] N. Sivapriya, M. K. Vanteru, K. K. Vaigandla, and G. Balakrishna, "Evaluation of PAPR, PSD, Spectral Efficiency, BER and SNR Performance of Multi-Carrier Modulation Schemes for 5G and Beyond," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 11, pp. 100–114, Nov. 2023, doi: 10.14445/23488379/IJEEE-V10I11P110.

[13] L. Ge, C. Shi, S. Niu, G. Chen, and Y. Guo, "Mixed RNN-DNN based channel prediction for massive MIMO-OFDM systems," *IET Communications*, vol. 17, no. 19, pp. 2152–2161, Dec. 2023, doi: 10.1049/cmu2.12685.

[14] H. Kasban, S. Hashima, S. Nassar, E. M. Mohamed, and M. A. M. El-Bendary, "Performance enhancing of MIMO-OFDM system utilizing different interleaving techniques with rate-less fountain raptor code," *IET Communications*, vol. 16, no. 20, pp. 2479–2491, 2022, doi: 10.1049/cmu2.12503.

[15] H. I. Bitat, F. Maamri, F. Khelfaoui, H. Djellab, Y. Belhocine, and Y. Messai, "Optimizing Spectral and Energy Efficiency of Massive MIMO Networks Using MVO and API Algorithms," *Journal of Telecommunications and Information Technology*, vol. 90, no. 1, pp. 81–91, 2025, doi: 10.26636/jtit.2025.1.1993.

[16] I. Jaiswal, R. G. Sangeetha, and M. Suchetha, "Performance of M-ary quadrature amplitude modulation -based orthogonal frequency division multiplexing for free space optical transmission," *IET Optoelectronics*, vol. 10, no. 4, pp. 156–162, Aug. 2016, doi: 10.1049/iet-opt.2015.0091.

[17] S. T. Başaran and G. K. Kurt, "Joint subcarrier and power allocation in OFDMA systems for outage minimization," *IEEE Communications Letters*, vol. 20, no. 10, pp. 2007–2010, Oct. 2016, doi: 10.1109/LCOMM.2016.2586038.

[18] S. Muy, D. Ron, and J. R. Lee, "Energy Efficiency Optimization for SWIPT-Based D2D-Underlaid Cellular Networks Using Multiagent Deep Reinforcement Learning," *IEEE Syst J*, vol. 16, no. 2, pp. 3130–3138, Jun. 2022, doi: 10.1109/JSYST.2021.3098860.

[19] R. Ali, I. Ashraf, A. K. Bashir, and Y. Bin Zikria, "Reinforcement-Learning-Enabled Massive Internet of Things for 6G Wireless Communications," *IEEE Communications Standards Magazine*, vol. 5, no. 2, pp. 126–131, Jun. 2021, doi: 10.1109/MCOMSTD.001.2000055.

[20] L. Liang, H. Ye, G. Yu, and G. Y. Li, "Deep-Learning-Based Wireless Resource Allocation with Application to Vehicular Networks," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 341–356, Feb. 2020, doi: 10.1109/JPROC.2019.2957798.

[21] J. Huang, Y. Yang, Z. Gao, D. He, and D. W. K. Ng, "Dynamic Spectrum Access for D2D-Enabled Internet of Things: A Deep Reinforcement Learning Approach," *IEEE Internet Things J*, vol. 9, no. 18, pp. 17793–17807, Sep. 2022, doi: 10.1109/JIOT.2022.3160197.

[22] X. Li, J. Fang, W. Cheng, H. Duan, Z. Chen, and H. Li, "Intelligent Power Control for Spectrum Sharing in Cognitive Radios: A Deep Reinforcement Learning Approach," *IEEE Access*, vol. 6, pp. 25463–25473, Apr. 2018, doi: 10.1109/ACCESS.2018.2831240.

[23] P. Lv, "Design and application of deep reinforcement learning algorithms based on unbiased exploration strategies for value functions," *Measurement: Sensors*, vol. 34, p. 101241, Aug. 2024, doi: 10.1016/j.measen.2024.101241.

[24] S. S. Omar, A. M. A. El-Haleem, I. I. Ibrahim, and A. M. Saleh, "Capacity Enhancement of Flying-IRS Assisted 6G THz Network Using Deep Reinforcement Learning," *IEEE Access*, vol. 11, pp. 101616–101629, 2023, doi: 10.1109/ACCESS.2023.3315660.

[25] G. Bacci, E. V. Belmega, P. Mertikopoulos, and L. Sanguinetti, "Energy-Aware Competitive Power Allocation for Heterogeneous Networks under QoS Constraints," *IEEE Trans Wirel Commun*, vol. 14, no. 9, pp. 4728–4742, Sep. 2015, doi: 10.1109/TWC.2015.2425397.