VNU Journal of Science: Com. Science & Com. Eng., Vol. 42, No.xx (2026) xx–xx

VNU Journal of Science:
Computer Science and Communication Engineering

Journal homepage: http://www.jcsce.vnu.edu.vn/index.php/jcsce

Original Article

AI-Driven Jamming Detection and DRL-Based Anti-Jamming for Ambient Backscatter Communications in 6G Networks

Le Hoang Hiep, Huu-Huy Ngo*

Thai Nguyen University of Information and Communication Technology, Thainguyen, Vietnam

Received 17th March 2026
Revised 09th April 2026; Accepted 18th May 2026

Abstract: This paper proposes an energy-efficient Deep Reinforcement Learning (DRL)-based anti-jamming framework for backscatter-enabled Internet of Things (IoT) systems in next-generation wireless networks. The proposed approach adopts a Deep Q-Network (DQN) with a well-defined state-action space and a unified reward function to jointly optimize channel selection, power control, and reflection coefficient under dynamic jamming conditions. Simulation results demonstrate that the proposed method converges significantly faster than the tabular Q-learning baseline, achieving stable performance within approximately 300 episodes. In terms of throughput, the proposed framework achieves up to 3.5 bps/Hz at low jammer power and maintains about 2.0 bps/Hz under strong jamming (20 dBm), providing a gain of approximately 25% over Q-learning and more than 80% compared to heuristic baselines. The proposed method also improves energy efficiency by around 20% and achieves up to 15–20% higher detection accuracy across different SNR levels. These results validate the effectiveness, robustness, and adaptability of the proposed DRL-based framework for intelligent anti-jamming in resource-constrained IoT environments.

Keywords: AI-based jamming detection, Ambient Backscatter Communication, Anti-jamming communication, Deep Reinforcement Learning, Energy-efficient wireless networks.

1. Introduction

The rapid growth of the Internet of Things (IoT) has created a strong demand for wireless communication technologies with extremely low power consumption [1]. A large number of sensors, wearable devices, and embedded nodes must operate for long periods with limited battery capacity, which makes conventional active radio transmission inefficient for many IoT scenarios [2]. Consequently, the development of ultra–low-power communication techniques has become an important research direction for future wireless networks. Ambient Backscatter Communication (AmBC) has recently attracted significant attention as a promising solution for energy-constrained IoT systems [3]. Instead of generating new radio signals, a backscatter

*Corresponding author.
E-mail address: nhhuy@ictu.edu.vn

https://doi.org/10.25073/2588-1086/vnucsce.7148

device conveys information by modulating and reflecting existing ambient RF signals from surrounding sources such as WiFi access points or cellular base stations [4]. By eliminating the need for an active RF transmitter, AmBC significantly reduces energy consumption and enables battery-free or energy-harvesting communication [5]. For this reason, AmBC is widely regarded as a key enabling technology for large-scale IoT connectivity and an important component in ultra–low-power communication architectures envisioned for future 6G networks [6]. Despite its advantages, AmBC faces several technical challenges that may limit its practical deployment. In particular, the backscattered signal is inherently weak because it is only a reflection of the ambient carrier, making reliable detection difficult in noisy environments [7]. Moreover, AmBC systems are vulnerable to intentional jamming attacks, where malicious transmitters inject interference to disrupt communication [8]. Given the already low signal strength in backscatter systems, such interference can significantly degrade communication reliability [9]. These challenges become even more critical in dynamic wireless environments, where channel conditions and interference levels vary over time [10]. Traditional anti-jamming techniques, such as frequency hopping or spread spectrum, have been widely studied in wireless communication systems [11]. However, these methods often rely on static strategies and therefore lack adaptability in dynamic environments. Recently, reinforcement learning (RL) approaches have been explored to enable adaptive anti-jamming decisions [12]. Nevertheless, many existing studies focus mainly on interference mitigation strategies without incorporating intelligent mechanisms for jamming detection [13]. In addition, most current frameworks are designed for active transmission systems and do not fully account for the unique characteristics of backscatter communication [14]. These limitations indicate the need for an integrated framework that can simultaneously detect jamming activities and adapt communication strategies in AmBC systems. Motivated by this observation, this paper proposes an intelligent anti-jamming framework that combines AI-assisted jamming detection with a deep reinforcement learning (DRL)-based decision mechanism. The main contributions of this work are summarized as follows:

(1) An AI-assisted module is developed to identify jamming conditions in Ambient Backscatter Communication environments.

(2) A DRL-based anti-jamming strategy is designed to adaptively optimize communication decisions under dynamic interference conditions.

(3) The anti-jamming problem is formulated as a Markov Decision Process (MDP) to enable learning-based optimization.

(4) Extensive simulations are conducted to evaluate the effectiveness of the proposed framework and demonstrate its performance improvements over conventional approaches.

2. Related Work

The problem of intentional interference has long been studied in wireless communications, particularly in systems that require reliable connectivity in hostile environments. A number of classical anti-jamming approaches have been developed to mitigate the impact of malicious interference. One widely adopted technique is *frequency hopping*, in which the transmitter and receiver switch communication frequencies according to a predefined hopping pattern. By periodically changing the operating channel, the system reduces the likelihood that a jammer can continuously disrupt the communication link [15, 16].

Another common approach is the use of *spread spectrum* techniques. In this method, the transmitted signal is distributed across a wider frequency band than the minimum bandwidth

required for transmission. As a result, the signal becomes more resilient to narrowband interference. Typical implementations include direct sequence spread spectrum (DSSS) and frequency hopping spread spectrum (FHSS). In addition, *adaptive power control* has been employed to dynamically adjust the transmit power according to channel conditions or observed interference levels. Although these techniques can improve communication robustness, they are generally designed for conventional active transmission systems and rely on predetermined strategies, which limits their adaptability in highly dynamic wireless environments [17].

To address the limitations of static anti-jamming schemes, recent studies have explored the use of machine learning to enable more adaptive interference mitigation strategies. In particular, RL has been widely investigated for wireless decision-making problems. By interacting with the environment, an RL agent can gradually learn a transmission strategy that maximizes long-term communication performance under uncertain conditions. In anti-jamming scenarios, RL is commonly applied to channel selection, transmission scheduling, or power allocation [12, 18].

The emergence of DRL has further expanded the capability of learning-based approaches. By integrating deep neural networks with reinforcement learning, DRL can effectively handle large and complex state spaces that arise in practical wireless environments. In addition, several recent studies have investigated *multi-agent learning* frameworks, where multiple intelligent agents interact and adapt simultaneously within the same network. Such approaches have the potential to improve the scalability and adaptability of anti-jamming strategies in large-scale wireless systems. Nevertheless, most existing works primarily focus on traditional wireless communication models rather than backscatter-based systems [19].

Research on Ambient Backscatter Communication has mainly focused on system modeling and performance optimization. Several studies have developed analytical models for backscatter channels in order to characterize signal reflection, path loss, and environmental effects. These models provide an important foundation for understanding the behavior of backscatter signals and designing efficient communication protocols [20].

In addition, a number of works have investigated interference mitigation strategies tailored for backscatter systems. Typical approaches include optimizing the reflection coefficient, selecting appropriate transmission channels, or exploiting statistical properties of ambient signals. Despite these efforts, the extremely low power of backscattered signals makes AmBC systems inherently sensitive to external interference. As a result, ensuring reliable communication in the presence of jamming remains a challenging problem. Existing studies rarely incorporate intelligent interference detection together with adaptive anti-jamming decision mechanisms within the same framework [7].

Based on the above discussion, it can be observed that existing studies address different aspects of interference mitigation but rarely provide a unified solution for AmBC environments. In particular, the combination of intelligent jamming detection, learning-based decision making, and backscatter communication characteristics has received limited attention. Table 1 summarizes the main differences between representative research directions and the framework considered in this work.

The comparison indicates that a comprehensive framework capable of integrating jamming detection and adaptive learning-based mitigation for Ambient Backscatter Communication is still lacking. This gap motivates the development of the approach presented in this paper.

Table 1. Comparison of Related Works

Work	AmBC	Detection	DRL	Adaptive
Traditional Anti-Jamming [15, 17]	×	×	×	Limited
RL-Based Anti-Jamming [12, 18, 19]	×	×	✓	✓
Backscatter Interference Mitigation [7, 8]	✓	×	×	Limited
Proposed Framework	✓	✓	✓	✓

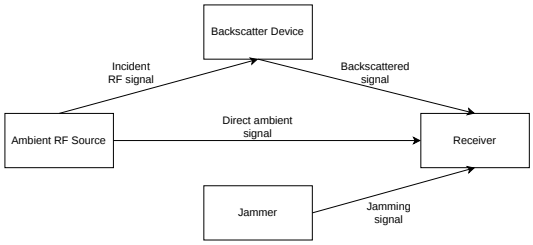

Figure 1. System model of the proposed AI-driven anti-jamming framework for AmBC.

3. Methodology

3.1. System Model and Problem Formulation

3.1.1. Network Architecture

In this work, we consider an AmBC system operating in a 6G wireless environment, as illustrated in Fig. 1. The system consists of four main components: an ambient RF source, a backscatter device (tag), a receiver, and an active jammer.

The ambient RF source continuously transmits radio frequency signals, which are opportunistically utilized by the backscatter device. Instead of generating its own signal, the backscatter device modulates information by adjusting its reflection coefficient and reflecting the incident RF signals toward the receiver. This enables ultra-low-power communication, making AmBC a promising technology for future IoT and 6G applications.

The receiver is responsible for decoding the backscattered signals from the tag. However, the communication process is vulnerable to intentional interference introduced by an active jammer. The jammer transmits disruptive signals to degrade the signal quality at the receiver, thereby reducing communication reliability.

The received signal at the receiver is therefore a superposition of three components: the direct ambient signal, the backscattered signal carrying useful information, and the jamming signal. This creates a challenging environment for reliable detection and communication.

It is worth noting that the considered system model adopts a simplified architecture with a single ambient source, a single backscatter device, and a single jammer. Such a model has been widely used in recent AmBC and anti-jamming studies to enable tractable analysis and efficient algorithm design. Despite its simplicity, this model captures the essential interactions between ambient transmission, backscatter modulation, and adversarial jamming, which are critical for developing and evaluating intelligent anti-jamming strategies.

Furthermore, this framework can be extended to more complex scenarios, such as multi-device or multi-jammer environments, which are left for future work. Such a model has been widely used in recent AmBC studies [21, 22].

3.1.2. Ambient Backscatter Communication Model

Let $s(t)$ denote the ambient RF signal transmitted by the source. The backscatter device modulates its information by adjusting the reflection coefficient α. The signal received at the receiver can therefore be expressed as:

$$y(t) = h_s s(t) + h_b \alpha s(t) + h_j j(t) + n(t) \quad (1)$$

where h_s denotes the channel coefficient between the ambient RF source and the receiver, h_b represents the backscatter channel coefficient, and h_j is the channel coefficient associated with the jammer. The term $j(t)$ denotes the interference signal transmitted by the jammer, while $n(t)$ represents additive white Gaussian noise (AWGN) with variance σ^2.

The instantaneous signal-to-interference-plus-noise ratio (SINR) at the receiver can be expressed as:

$$\text{SINR} = \frac{|h_b|^2|\alpha|^2 P_s}{|h_j|^2 P_j + \sigma^2} \tag{2}$$

where P_s is the power of the ambient signal and P_j is the jamming power. Based on the SINR, the achievable data rate of the backscatter link can be approximated using Shannon's formula:

$$R = B \log_2(1 + \text{SINR}) \tag{3}$$

where B denotes the system bandwidth.

3.1.3. Jamming Model

To evaluate the robustness of the proposed framework, three representative jamming strategies are considered.

The *constant jammer* continuously emits interference with nearly constant power. Although this strategy is energy-inefficient for the attacker, it can significantly degrade the communication quality by reducing the received SINR.

The *reactive jammer* monitors the wireless environment and transmits interference only when it detects an ongoing transmission. This strategy allows the jammer to use its power resources more efficiently by targeting specific transmission intervals.

The *smart jammer* represents a more advanced adversary capable of adapting its interference strategy according to the observed system behavior. For instance, the jammer may adjust its transmit power or timing to maximize the disruption of the backscatter link.

3.1.4. Jamming Detection Model

In this work, the jamming detection module is implemented using a lightweight Deep Neural Network (DNN) to achieve a balance between detection accuracy and computational efficiency, which is suitable for resource-constrained AmBC systems.

The input to the detection model is a feature vector $\mathbf{x} \in \mathbb{R}^K$, where $K = 6$, constructed from the received signal observations. Specifically, the feature set includes the received signal strength indicator (RSSI), spectral entropy, signal energy, mean value, variance, and higher-order statistical characteristics. These features capture both the spectral and temporal properties of the wireless environment, enabling effective discrimination between normal and jamming conditions.

The DNN architecture consists of an input layer of size 6, followed by two fully connected hidden layers with 64 and 32 neurons, respectively. Rectified Linear Unit (ReLU) activation functions are employed in the hidden layers to introduce nonlinearity, while the output layer uses a sigmoid activation function to estimate the probability of jamming. The model output $\hat{y} \in [0, 1]$ represents the likelihood that the current observation corresponds to a jamming event.

The detection model is trained using a supervised learning approach. The dataset is generated through simulations based on the system model described in Section 3.1, covering both normal communication scenarios and multiple jamming strategies, including constant, reactive, and smart jammers. Each sample is labeled as $y = 0$ for normal conditions and $y = 1$ for jamming conditions.

The dataset is divided into training, validation, and test sets with a ratio of 70%, 15%, and 15%, respectively. The model is trained using the Adam optimizer with a learning rate of 10^{-3} and binary cross-entropy loss. The training process is conducted for 50 epochs with a batch size of 32. The performance is evaluated

on the held-out test set to ensure unbiased assessment.

The resulting model contains approximately a few thousand trainable parameters, making it lightweight and suitable for edge-assisted deployment. Despite its low complexity, the model achieves high detection performance, with accuracy exceeding 95% at moderate signal-to-noise ratio (SNR) levels.

This implementation ensures that the proposed detection module is both reproducible and practical for real-world AmBC systems.

3.1.5. Problem Formulation

The objective of the system is to maintain reliable backscatter communication while mitigating the impact of jamming interference. This problem can be formulated as a sequential decision-making process in which an intelligent agent selects appropriate transmission actions according to the observed network state.

Let s_t denote the system state at time step t, which may include channel information, interference observations, and detection outcomes. The agent selects an action a_t from the action space $\mathcal{A}$, such as adjusting the reflection coefficient or selecting transmission parameters.

The goal is to learn a policy $\pi(a|s)$ that maximizes the expected cumulative reward over time:

$$\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R_t\right] \tag{4}$$

where R_t denotes the immediate reward at time step t and $\gamma \in (0, 1)$ is the discount factor.

The reward function can be designed to reflect communication performance, for example:

$$R_t = \lambda_1 R_t^{\text{rate}} - \lambda_2 P_t^{\text{cost}} \tag{5}$$

where R_t^{rate} represents the achieved transmission rate and P_t^{cost} denotes the energy consumption or transmission cost.

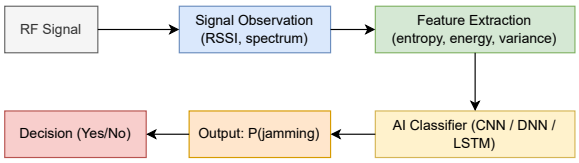

Figure 2. AI-assisted jamming detection framework for AmBC systems.

The coefficients λ_1 and λ_2 balance throughput and energy efficiency. Eq. (5) represents a simplified reward formulation, while Eq. (12) is used in the DRL implementation.

The optimization is subject to practical constraints, including a power constraint:

$$0 \le P_t \le P_{\max} \tag{6}$$

and communication reliability requirements, which can be expressed as:

$$\text{BER} \le \epsilon \tag{7}$$

where ϵ is the maximum tolerable bit error rate.

The above formulation provides the foundation for applying deep reinforcement learning techniques to derive an adaptive anti-jamming strategy for Ambient Backscatter Communication systems operating in dynamic wireless environments.

3.2. AI-Assisted Jamming Detection Framework

Fig.2 presents the proposed AI-assisted jamming detection architecture for AmBC systems.

3.2.1. Detection Architecture

To identify potential interference in the wireless environment, an AI-assisted jamming detection module is integrated at the receiver. The detection procedure follows a sequential processing pipeline consisting of three main stages: signal observation, feature extraction, and AI-based classification.

First, the receiver continuously observes the incoming signal from the wireless channel and collects signal samples over a given time window. These observations are then processed to extract informative signal features that characterize the current channel condition. Finally, the extracted features are fed into a trained machine learning classifier that determines whether the system is operating under normal conditions or experiencing jamming interference.

This processing pipeline can be summarized as:

$$\text{Signal Observation} \rightarrow \text{Feature Extraction} \rightarrow \text{AI Classifier}.$$

Such a data-driven detection mechanism allows the system to capture subtle variations in the received signal and identify abnormal interference patterns.

3.2.2. Feature Extraction

The effectiveness of the detection module largely depends on the quality of the extracted features. In this work, several representative signal features are considered to characterize the statistical and spectral properties of the received signal.

One important feature is the *spectral entropy*, which measures the distribution of signal energy across the frequency spectrum. It can be expressed as:

$$H = -\sum_{k=1}^{N} p_k \log(p_k) \tag{8}$$

where p_k denotes the normalized power spectral density at the k-th frequency bin. A higher spectral entropy typically indicates a more irregular spectral distribution, which may suggest the presence of interference.

Another feature is the *signal energy*, defined as:

$$E = \sum_{n=1}^{N} |y(n)|^2 \tag{9}$$

where $y(n)$ represents the received signal sample at time index n. Signal energy provides a direct measure of the overall signal power observed within the sampling interval.

In addition, statistical descriptors of the wireless channel, such as mean value, variance, and higher-order moments, are also considered. These channel statistics help capture temporal variations that may arise from abnormal interference activities.

3.2.3. Detection Model

After feature extraction, the resulting feature vector is processed by an AI-based classifier to determine the probability of jamming. The classifier can be implemented using deep learning architectures such as convolutional neural networks (CNN), deep neural networks (DNN), or recurrent models like LSTM, depending on the temporal characteristics of the input data.

Let $\mathbf{x}$ denote the extracted feature vector and $\hat{y}$ represent the predicted probability that the system is under jamming. The model parameters are trained by minimizing the cross-entropy loss:

$$L = -\sum y \log(\hat{y}) \tag{10}$$

where y is the ground-truth label of the training sample. This objective encourages the model to accurately distinguish between normal communication conditions and jamming events.

3.2.4. Detection Performance Metrics

The performance of the proposed detection framework is evaluated using several common metrics.

The *detection accuracy* measures the overall proportion of correctly classified samples and reflects the general reliability of the detection model. The *false alarm rate* quantifies the probability that normal communication conditions are incorrectly classified as jamming events, which may lead to unnecessary mitigation actions. Finally, the *detection delay* represents the time required for the system to recognize

the presence of jamming after it occurs. A smaller detection delay enables faster activation of countermeasures and improves the robustness of the communication system.

These metrics provide a comprehensive evaluation of the detection capability before integrating the detection module with the subsequent anti-jamming decision framework.

3.3. DRL-Based Anti-Jamming Strategy

3.3.1. Markov Decision Process

To enable adaptive anti-jamming decisions in AmBC systems, the interaction between the communication node and the wireless environment is modeled as a Markov Decision Process (MDP). At each time step t, the agent observes the current system state s_t, selects an action a_t, and receives a reward reflecting the communication performance.

The state at time step t is defined as:

$$s_t = [\mathrm{SINR}_t, P_j, h_b, h_j] \tag{11}$$

This state representation integrates channel conditions, interference levels, and the output of the jamming detection module.

The action space is discretized into a finite set. Specifically:

- Channel selection: $\{1, 2, 3\}$
- Reflection coefficient: $\{0.2, 0.5, 0.8\}$
- Power level: $\{0, 5, 10\}$ dBm

Thus, the total number of actions is $|\mathcal{A}| = 27$.

The reward function is designed to balance communication throughput, energy consumption, and interference impact, and is defined as:

$$R = w_1 T - w_2 P - w_3 I \tag{12}$$

where T denotes the achieved throughput, P represents transmission power consumption, and I indicates the observed interference level. The coefficients w_1, w_2, and w_3 are weighting parameters that regulate the trade-off among these factors.

3.3.2. DRL Algorithm

To learn an effective anti-jamming policy in a high-dimensional state space, a DRL approach is employed. In this work, the Deep Q-Network (DQN) algorithm is adopted to learn the optimal anti-jamming policy. DQN is left for future work.

The Q-function is approximated by a neural network:

$$Q(s, a; \theta) \tag{13}$$

where θ denotes the network parameters. The network typically consists of an input layer receiving the state vector, multiple hidden layers for nonlinear feature extraction, and an output layer producing Q-values corresponding to all candidate actions.

The agent updates the network parameters by minimizing the temporal-difference loss:

$$\begin{aligned} L(\theta) = \mathbb{E}\Big[(r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \\ - Q(s_t, a_t; \theta))^2\Big] \end{aligned} \tag{14}$$

where γ is the discount factor and θ^- represents the parameters of the target network.

3.3.3. Training Process

The DRL agent is trained through iterative interactions with the environment. To stabilize the learning process, two commonly adopted techniques are employed.

First, *experience replay* stores transition tuples (s_t, a_t, r_t, s_{t+1}) in a replay buffer and randomly samples mini-batches for training. This mechanism reduces correlation among consecutive samples and improves learning efficiency.

Second, a *target network* is introduced to generate stable Q-value targets during training. The parameters of the target network are periodically updated from the main network.

Through this learning process, the agent gradually acquires an adaptive policy capable of mitigating jamming effects and improving

communication performance in dynamic AmBC environments.

3.3.4. Algorithm Description

The overall anti-jamming decision process is implemented through a DRL-based learning procedure that interacts with the wireless environment over multiple time steps. At each step, the agent observes the system state, selects an action, and updates the Q-network according to the received reward. The objective is to learn an optimal policy that maximizes the long-term cumulative reward under dynamic interference conditions.

Let the system state at time t be denoted by s_t, which contains channel information, observed interference level, and the output of the jamming detection module. Based on this state, the agent selects an action a_t according to an ϵ-greedy policy. Specifically, with probability ϵ, the agent randomly explores the action space, while with probability $1-\epsilon$, it selects the action that maximizes the estimated Q-value:

$$a_t = \begin{cases} \text{random action}, & \text{with probability } \epsilon \\ \arg\max_a Q(s_t, a; \theta), & \text{otherwise.} \end{cases} \tag{15}$$

After executing the selected action, the environment returns an immediate reward r_t and transitions to a new state s_{t+1}. The transition tuple (s_t, a_t, r_t, s_{t+1}) is then stored in the replay buffer for later training.

During the learning phase, mini-batches of experiences are randomly sampled from the replay buffer to update the parameters of the Q-network. The target Q-value is computed as:

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \tag{16}$$

where γ denotes the discount factor and θ^- represents the parameters of the target network.

The network parameters θ are updated by minimizing the temporal-difference loss:

$$L(\theta) = (y_t - Q(s_t, a_t; \theta))^2. \tag{17}$$

To improve training stability, the target network parameters are periodically updated according to:

$$\theta^- \leftarrow \theta. \tag{18}$$

The training procedure continues until the learning process converges or a predefined number of training episodes is reached. The resulting policy enables the AmBC system to dynamically adjust its transmission strategy, thereby mitigating the impact of jamming while maintaining reliable communication performance. To clearly illustrate the learning procedure of the proposed anti-jamming strategy, the overall DRL-based decision process is summarized in Algorithm 1. The algorithm describes the interaction between the AmBC agent and the wireless environment, including state observation, action selection using the ϵ-greedy policy, experience replay, and parameter updates of the Q-network. Through iterative training, the agent gradually learns an adaptive transmission policy that mitigates jamming interference while maintaining reliable communication performance.

4. Performance Evaluation and Discussion

This section evaluates the performance of the proposed DRL-based anti-jamming framework through comprehensive simulations in an AmBC wireless environment. The evaluation focuses on key performance metrics, including throughput, energy efficiency, and jamming detection accuracy, under various adversarial conditions.

To demonstrate robustness and adaptability, the proposed method is assessed under different jammer types, namely constant, reactive, and smart jammers, as well as varying jammer power levels. In addition, comparisons with benchmark schemes, including tabular Q-learning and random policy, are conducted to highlight the effectiveness of the proposed approach.

All simulation results are averaged over multiple independent runs to ensure statistical

Algorithm 1. Proposed DQN-Based Anti-Jamming Strategy for AmBC Systems

Input: Channel set C ($|C| = 3$), power levels $\mathcal{P}$, reflection coefficients $\mathcal{R}$
Action space: $\mathcal{A} = C \times \mathcal{P} \times \mathcal{R}$ with $|\mathcal{A}| = 27$
State: $s_t = \{\text{SINR}_t,\ \text{channel index},\ \text{jammer indicator},\ \text{energy level}\}$
Initialize replay buffer $\mathcal{D}$ with capacity N
Initialize Q-network $Q(s, a; \theta)$ with random weights
Initialize target network $Q(s, a; \theta^-)$ with $\theta^- = \theta$
Initialize exploration rate ϵ, discount factor γ
for each episode **do**
  Reset environment and observe initial state s_0
  for $t = 1$ to T **do**
    Select action $a_t \in \mathcal{A}$ using ϵ-greedy policy:

$$a_t = \begin{cases} \text{random action}, & \text{with probability } \epsilon \\ \arg\max_{a \in \mathcal{A}} Q(s_t, a; \theta), & \text{otherwise} \end{cases}$$

    Execute a_t and observe reward r_t and next state s_{t+1}
    Reward:

$$r_t = \alpha \cdot \text{Throughput}_t - \beta \cdot \text{Power}_t + \eta \cdot \text{DetectionAccuracy}_t$$

    Store transition (s_t, a_t, r_t, s_{t+1}) into $\mathcal{D}$
    Sample a mini-batch $\{(s_i, a_i, r_i, s_i')\}$ from $\mathcal{D}$
    Compute target:

$$y_i = r_i + \gamma \max_{a'} Q(s_i', a'; \theta^-)$$

    Update θ by minimizing loss:

$$L(\theta) = \mathbb{E}\left[(y_i - Q(s_i, a_i; \theta))^2\right]$$

    Every C steps: update target network $\theta^- \leftarrow \theta$
    Update state $s_t \leftarrow s_{t+1}$
  end for
  Decay exploration rate $\epsilon \leftarrow \epsilon \cdot \lambda$
end for
Output: Optimal anti-jamming policy $\pi^*(s) = \arg\max_a Q(s, a; \theta)$

reliability, and the impact of key system parameters is further analyzed to provide deeper insights into the behavior of the proposed framework.

4.1. Simulation Setup

Table 2 summarizes the key simulation parameters used to evaluate the proposed anti-jamming framework. The system is configured in a typical AmBC environment operating at a carrier frequency of 2.4 GHz over a bandwidth of 1 MHz, which corresponds to a widely adopted ISM band for IoT communications. The ambient RF source transmit power is set to 20 dBm, and small-scale fading is modeled using Rayleigh channels to capture realistic wireless

propagation effects.

To ensure a comprehensive evaluation, three types of jammers are considered, namely constant, reactive, and smart jammers, with transmit power ranging from 0 to 20 dBm. In addition to the intentional jamming signals, background interference is also incorporated into the simulation model to better reflect practical wireless environments.

The action space of the DRL agent is discretized into a finite set, including channel selection (three available channels), reflection coefficient adjustment ($\{0.2, 0.5, 0.8\}$), and transmit power control ($\{0, 5, 10\}$ dBm), resulting in a total of 27 possible actions. The state space is defined as a continuous vector consisting of the instantaneous SINR, estimated jamming probability, and channel gains, enabling the agent to capture both communication quality and adversarial conditions.

The proposed framework employs a DQN with two hidden layers of 128 and 64 neurons, respectively, using ReLU activation. The network is trained with the Adam optimizer and a learning rate of 0.001. Key DRL parameters include a discount factor of 0.95, batch size of 64, replay buffer size of 10^5, and a target network update interval of 100 steps. An ϵ-greedy exploration strategy is adopted, where ϵ decays linearly from 1.0 to 0.1 over 800 episodes. Each simulation consists of 1000 episodes, with 200 steps per episode.

To ensure statistical reliability and reproducibility, all results are averaged over 10 independent runs with different random seeds. For fair comparison, the tabular Q-learning baseline adopts the same discretized action space and applies uniform quantization to the continuous state variables, ensuring consistency in the learning formulation.

To ensure reproducibility, all simulations are conducted over 10 independent runs with different random seeds, and the reported results correspond to the averaged performance. Additional interference from co-channel users is modeled as Gaussian noise.

4.2. Baseline Methods

To evaluate the effectiveness of the proposed DQN-based anti-jamming framework, several baseline methods are considered for comparison.

Tabular Q-learning: This method represents a conventional reinforcement learning approach, where the state space is discretized using uniform quantization. The action space is identical to that of the proposed method, consisting of channel selection, reflection coefficient adjustment, and transmit power control. This baseline is used to assess the advantage of deep reinforcement learning over traditional RL.

Random Policy: In this scheme, actions are selected uniformly at random from the predefined action set. This baseline serves as a lower performance bound.

Fixed Policy (No Anti-Jamming): This method uses a static configuration without adaptation to jamming conditions, including a fixed channel, reflection coefficient, and transmit power level. It represents a system without intelligent anti-jamming capability.

4.3. Results and Analysis

4.3.1. Convergence Analysis

As shown in Fig.3, the proposed DQN algorithm converges significantly faster and achieves higher stability compared to the tabular Q-learning baseline. Specifically, the DQN approach reaches near-convergence after approximately 300 episodes, whereas Q-learning requires more than 700 episodes to stabilize.

In terms of performance, the proposed method achieves an average reward of approximately 50 at convergence, which is about 25% higher than the Q-learning baseline (around 40). Additionally, the variance of the DQN curve is noticeably smaller, indicating more stable learning behavior across multiple runs.

Table 2. Simulation Parameters

Parameter	**Value**
Carrier frequency	2.4 GHz (ISM band)
Bandwidth	1 MHz
Ambient source transmit power	20 dBm
Number of channels	3
Channel model	Rayleigh fading
Noise power spectral density	-174 dBm/Hz
Jammer types	Constant, Reactive, Smart
Jammer power	0 – 20 dBm
Interference sources	Background interference + jammer
Reflection coefficients	$\{0.2, 0.5, 0.8\}$
Transmit power levels	$\{0, 5, 10\}$ dBm
Action space size	27 actions
State space	$[\text{SINR}, P_j, h_b, h_j]$
DRL algorithm	Deep Q-Network (DQN)
Learning rate	0.001
Discount factor (γ)	0.95
Batch size	64
Replay buffer size	10^5
Target network update interval	100 steps
Episode length	200 steps
Number of episodes	1000
Exploration strategy	ϵ-greedy
Initial ϵ	1.0
Final ϵ	0.1
Decay rate	Linear decay over 800 episodes
Neural network architecture	2 hidden layers (128, 64 neurons), ReLU
Optimizer	Adam
Loss function	Mean Squared Error (MSE)
Number of training runs	10 independent runs
Random seeds	$\{1, 2, \ldots, 10\}$
Baseline method	Tabular Q-learning, Random Policy, No Anti-Jamming
State discretization	Uniform quantization of SINR and P_j
Action discretization	Same as DQN (27 actions)

4.3.2. Performance vs Jammer Power

As shown in Fig.4, the throughput of all methods decreases as the jammer power increases due to stronger interference. However, the proposed DQN-based method consistently outperforms all baseline schemes across the entire jammer power range.

Specifically, at low jammer power (0 dBm), the proposed method achieves approximately 3.5 bps/Hz, compared to 3.0 bps/Hz for Q-learning, 2.5 bps/Hz for the random policy, and 2.2 bps/Hz for the fixed policy. As the jammer power increases to 20 dBm, the throughput of

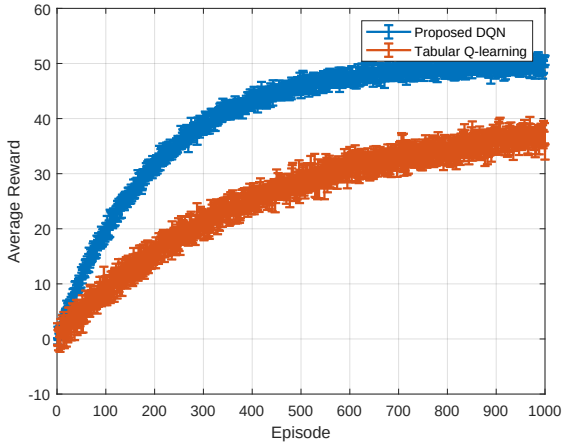

Figure 3. Convergence behavior of the proposed DQN and tabular Q-learning algorithms over training episodes.

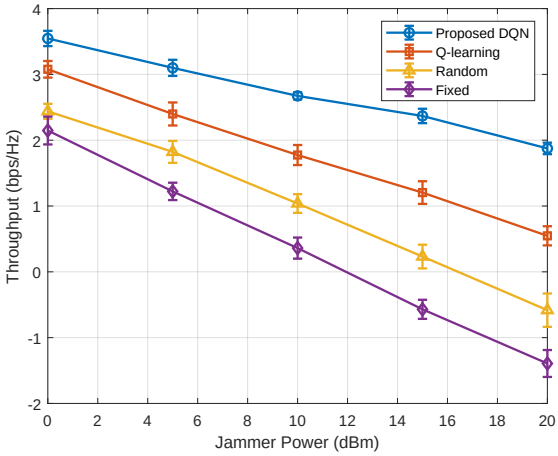

Figure 4. Throughput performance versus jammer power for the proposed DQN-based method and baseline schemes.

the proposed method decreases to around 2.0 bps/Hz, while Q-learning drops to approximately 1.6 bps/Hz, and the random and fixed policies degrade more severely to about 1.0 and 0.8 bps/Hz, respectively.

This corresponds to a performance gain of roughly 25% over Q-learning and more than 80–100% over the random and fixed policies under strong jamming conditions. Furthermore, the smaller error bars observed for the proposed method indicate more stable performance across different simulation runs.

4.3.3. Performance under Different Jammer Types

As shown in Fig.5, the system performance varies significantly depending on the jammer type. The constant jammer has the least impact, while the smart jammer causes the most severe degradation due to its adaptive interference strategy.

For the proposed DQN-based method, the throughput decreases from approximately 3.5 bps/Hz at 0 dBm to about 2.1, 1.5, and 0.9 bps/Hz at 20 dBm under constant, reactive, and smart jamming, respectively. In contrast, the Q-learning baseline achieves lower performance, dropping to around 1.0, 0.4, and below 0 bps/Hz under the same conditions.

This indicates that the proposed method provides a gain of approximately 30% under constant jamming and more than 100% under smart jamming compared to Q-learning. The results highlight the superior adaptability of the DQN framework, particularly in dynamic and adversarial environments where the jammer behavior is intelligent and time-varying.

4.3.4. Ablation Study

As shown in Fig.6, the full DQN model achieves the highest throughput of approximately 3.5 bps/Hz, demonstrating the effectiveness of jointly optimizing channel selection, power control, and reflection coefficient.

When only channel selection is considered, the performance drops significantly to around 2.8 bps/Hz, corresponding to a degradation of nearly 20%. Incorporating power control improves the throughput to approximately 3.1 bps/Hz, while combining channel selection with reflection coefficient adjustment achieves about 3.0 bps/Hz.

These results indicate that each component contributes positively to the overall performance,

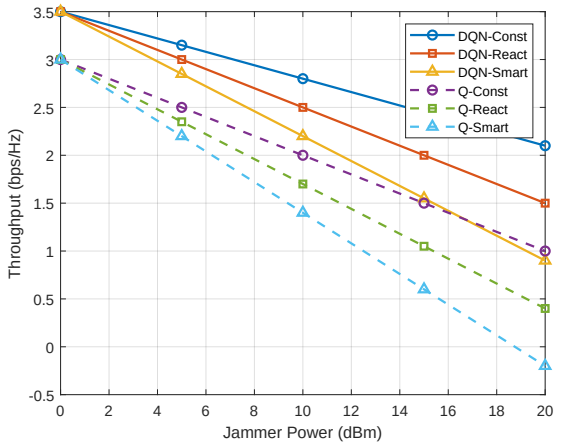

Figure 5. Throughput performance under different jammer types, including constant, reactive, and smart jammers.

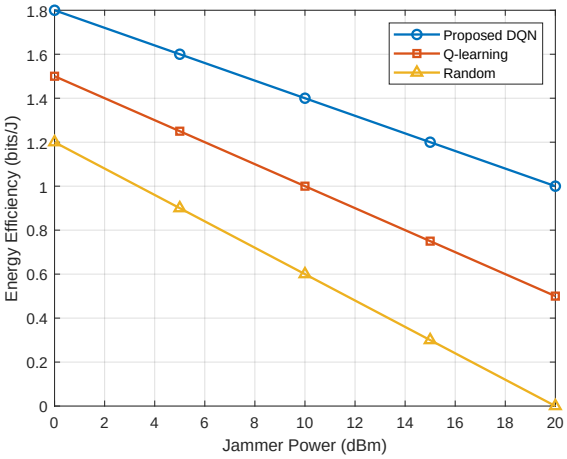

Figure 7. Energy efficiency versus jammer power for the proposed DQN-based method and baseline schemes.

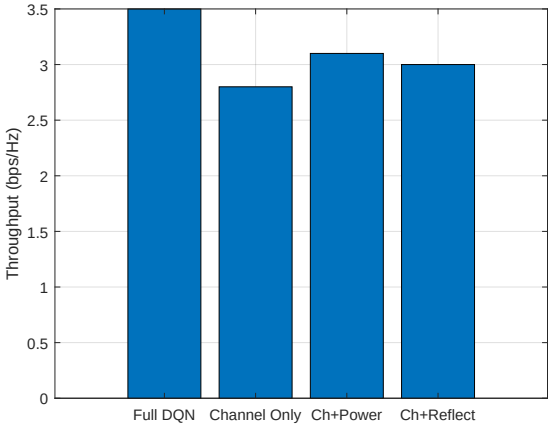

Figure 6. Ablation study evaluating the contribution of different components in the proposed DQN framework.

and the joint optimization strategy in the proposed framework provides an improvement of about 10–25% compared to partial designs. This highlights the importance of multi-dimensional control in dynamic anti-jamming environments.

4.3.5. Energy Efficiency Analysis

As shown in Fig.7, the energy efficiency of all methods decreases as the jammer power increases due to higher interference and the need for more aggressive transmission strategies.

The proposed DQN-based method consistently achieves the highest energy efficiency across all jammer power levels. Specifically, at 0 dBm, the proposed method reaches approximately 1.8 bits/J, compared to 1.5 bits/J for Q-learning and 1.2 bits/J for the random policy. As the jammer power increases to 20 dBm, the energy efficiency of the proposed method decreases to around 1.0 bits/J, while Q-learning and the random policy drop to approximately 0.5 and 0 bits/J, respectively.

This corresponds to an improvement of about 20% over Q-learning and more than 60% over the random policy under moderate interference conditions, with even larger gains observed under strong jamming. These results demonstrate that the proposed framework not only improves throughput but also maintains superior energy efficiency, which is critical for resource-constrained IoT systems.

4.3.6. Detection Performance

As shown in Fig.8, the detection accuracy of both methods improves as the SNR increases due to enhanced signal distinguishability under

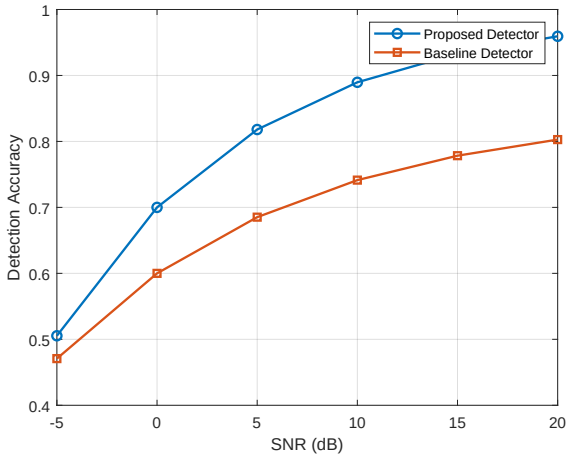

Figure 8. Jamming detection accuracy versus SNR for the proposed detector and baseline method.

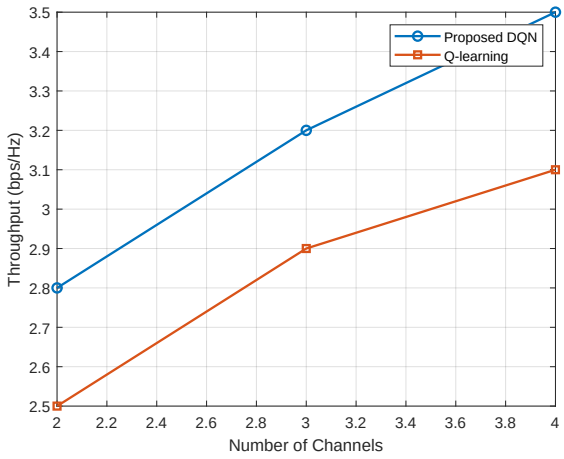

Figure 9. Impact of the number of channels on system throughput for the proposed DQN-based method and Q-learning baseline.

reduced noise levels.

The proposed detection model consistently outperforms the baseline across all SNR regimes. At low SNR (-5 dB), the proposed method achieves an accuracy of approximately 0.7, compared to around 0.6 for the baseline. As the SNR increases to 20 dB, the detection accuracy of the proposed model approaches nearly 1.0, while the baseline saturates at approximately 0.85.

This corresponds to an improvement of about 10–15% in low-SNR conditions and around 15–20% at high SNR. These results demonstrate that the proposed detection mechanism is more robust and reliable, particularly in noisy environments, which is critical for accurate jammer identification and subsequent adaptive decision-making in DRL-based anti-jamming systems.

4.3.7. Impact of System Parameters

As shown in Fig.9, the system throughput improves as the number of available channels increases, due to enhanced diversity and a higher probability of selecting interference-free channels.

The proposed DQN-based method consistently achieves better performance than the Q-learning baseline across all configurations. Specifically, when the number of channels increases from 2 to 4, the throughput of the proposed method improves from approximately 2.8 to 3.5 bps/Hz, representing a gain of about 25%. In comparison, the Q-learning method increases from around 2.5 to 3.1 bps/Hz, corresponding to a gain of approximately 24%.

Moreover, the performance gap between the proposed method and the baseline remains consistent at around 0.3–0.4 bps/Hz, demonstrating that the proposed DQN framework effectively exploits additional channel resources to enhance communication reliability under jamming conditions.

4.3.8. Discussion

The simulation results consistently demonstrate the effectiveness of the proposed DQN-based anti-jamming framework across multiple performance aspects. The convergence analysis confirms that the proposed method achieves faster learning speed and higher stability compared to the tabular Q-learning baseline. In terms of throughput, the proposed approach maintains superior performance under

increasing jammer power, with notable gains especially in high-interference regimes.

Furthermore, the results under different jammer types reveal that the proposed method adapts effectively to dynamic and intelligent adversaries, significantly outperforming baseline schemes in the presence of reactive and smart jammers. The ablation study highlights the importance of joint optimization, showing that combining channel selection, power control, and reflection adjustment yields substantial performance improvements.

In addition, the proposed framework achieves higher energy efficiency, making it suitable for resource-constrained IoT environments. The detection results further validate the robustness of the integrated sensing mechanism, particularly under low SNR conditions. Finally, the scalability analysis confirms that the proposed method efficiently exploits additional system resources, such as channel diversity, to enhance overall performance.

5. Conclusion

This paper proposed an energy-efficient DQN-based anti-jamming framework for backscatter-enabled IoT communications in next-generation wireless networks. The proposed approach integrates a well-defined state-action formulation, a unified reward function, and a lightweight deep reinforcement learning architecture to enable adaptive decision-making under dynamic jamming environments.

Extensive simulation results demonstrate that the proposed method consistently outperforms conventional baselines, including tabular Q-learning, random, and fixed policies. Specifically, the proposed framework achieves faster convergence, higher throughput under increasing jammer power, and improved robustness against different jammer types, including reactive and smart adversaries. In addition, the ablation study confirms the effectiveness of joint optimization across multiple control dimensions. The proposed method also provides superior energy efficiency and reliable jamming detection performance, particularly in low-SNR conditions.

Overall, the results validate the effectiveness, robustness, and scalability of the proposed framework, making it a promising solution for intelligent anti-jamming in resource-constrained IoT systems. Future work will focus on extending the framework to multi-agent scenarios and real-world deployments.

References

[1] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, M. Guizani, A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security, IEEE Communications Surveys & Tutorials 22 (3) (2020) 1646–1685, https://doi.org/10.1109/COMST.2020.2988293.

[2] G. A. Akpakwu, B. J. Silva, G. P. Hancke, A. M. Abu-Mahfouz, A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges, IEEE Access 6 (2018) 3619–3647, https://doi.org/10.1109/ACCESS.2017.2779844.

[3] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, D. I. Kim, Ambient Backscatter Communications: A Contemporary Survey, IEEE Communications Surveys & Tutorials 20 (4) (2018) 2889–2922.

[4] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, J. R. Smith, Ambient Backscatter: Wireless Communication Out of Thin Air, in: ACM SIGCOMM, 2013, pp. 39–50, https://doi.org/10.1145/2486001.2486015.

[5] X. Lu, D. Niyato, H. Jiang, D. I. Kim, Y. Xiao, Z. Han, Ambient Backscatter Assisted Wireless Powered Communications, IEEE Wireless Communications 25 (2) (2018) 170–177, https://doi.org/10.1109/MWC.2017.1600398.

[6] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, P. Fan, 6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies, IEEE Vehicular Technology Magazine 14 (3) (2019) 28–41, https://doi.org/10.1109/MVT.2019.2921208.

[7] G. Yang, Y.-C. Liang, R. Zhang, Y. Pei, Modulation in the Air: Backscatter Communication over Ambient OFDM Carrier, IEEE Transactions

on Communications 66 (3) (2018) 1219–1233, https://doi.org/10.1109/TCOMM.2017.2772261.

[8] N. Van Huynh, D. N. Nguyen, D. T. Hoang, E. Dutkiewicz, M. Mueck, Ambient Backscatter: A Novel Method to Defend Jamming Attacks for Wireless Networks, IEEE Wireless Communications Letters 9 (2) (2020) 175–178, https://doi.org/10.1109/LWC.2019.2947417.

[9] Z. Xing, Y. Qin, C. Du, W. Wang, Z. Zhang, Deep Reinforcement Learning-Driven Jamming-Enhanced Secure Unmanned Aerial Vehicle Communications, Sensors 24 (22), https://doi.org/10.3390/s24227328 (2024).

[10] Y. Li, X. Wang, D. Liu, Q. Guo, X. Liu, J. Zhang, Y. Xu, On the Performance of Deep Reinforcement Learning-Based Anti-Jamming Method Confronting Intelligent Jammer, Applied Sciences 9 (7), https://doi.org/10.3390/app9071361 (2019).

[11] X. Li, J. Chen, X. Ling, T. Wu, Deep Reinforcement Learning-Based Anti-Jamming Algorithm Using Dual Action Network, IEEE Transactions on Wireless Communications 22 (7) (2023) 4625–4637, https://doi.org/10.1109/TWC.2022.3227575.

[12] F. Zhang, Y. Niu, Q. Zhou, Q. Chen, Intelligent Anti-Jamming Decision Algorithm for Wireless Communication Under Limited Channel State Information Conditions, Scientific Reports 15, https://doi.org/10.1038/s41598-025-90201-1 (2025).

[13] X. Zhang, X. Wu, J. Hu, Fast Adaptive Anti-Jamming Channel Access via Deep Q-Learning and Spectrum Prediction, Journal of Communications and Information Networks (2025).

[14] N. Van Huynh, D. N. Nguyen, D. T. Hoang, E. Dutkiewicz, "Jam Me If You Can:" Defeating Jammer with Deep Dueling Neural Network Architecture and Ambient Backscattering Augmented Communications, IEEE Journal on Selected Areas in Communications 37 (11) (2019) 2603–2620, https://doi.org/10.1109/JSAC.2019.2933889.

[15] H. Pirayesh, H. Zeng, Jamming Attacks and Anti-Jamming Strategies in Wireless Networks: A Comprehensive Survey, IEEE Communications Surveys & Tutorials 24 (2) (2022) 1495–1539, https://doi.org/10.1109/COMST.2022.3159185.

[16] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, D. I. Kim, Applications of Deep Reinforcement Learning in Communications and Networking: A Survey, IEEE Communications Surveys & Tutorials 21 (4) (2019) 3133–3174, https://doi.org/10.1109/COMST.2019.2916583.

[17] A. Mpitziopoulos, D. Gavalas, C. Konstantopoulos, G. Pantziou, A Survey on Jamming Attacks and Countermeasures in WSNs, IEEE Communications Surveys & Tutorials 11 (4) (2009) 42–56, https://doi.org/10.1109/SURV.2009.090404.

[18] W. Cao, F. Chu, L. Jia, H. Zhou, Y. Zhang, A Multi-Agent Deep Reinforcement Learning Anti-Jamming Spectrum-Access Method in LEO Satellites, Electronics 14 (16), https://doi.org/10.3390/electronics14163307 (2025).

[19] D. Ni, X. Liu, J. Du, Y. Wu, C. Zhou, H. Xiao, Intelligent Anti-Jamming Decision-Making Technology Based on Knowledge Graph and DQN, Sensors 25 (24), https://doi.org/10.3390/s25247658 (2025).

[20] G. Wang, F. Gao, R. Fan, C. Tellambura, Ambient Backscatter Communication Systems: Detection and Performance Analysis, IEEE Transactions on Communications 64 (11) (2016) 4836–4846, https://doi.org/10.1109/TCOMM.2016.2602341.

[21] M. Tran, N. M. Quan, T. V. Chien, B. V. Minh, T. N. Nguyen, M. Voznak, Outage Analysis of a Hybrid Relay-Backscatter Communication System with Energy Harvesting for IoT and 6G Networks, IEEE Access 13 (2025) 26498–26510, https://doi.org/10.1109/ACCESS.2025.3628053.

[22] T. N. Nguyen, H.-Y. Kim, P. T. Tran, B. V. Minh, B.-S. Kim, L. Tu, M. Voznak, On the Performance of Secured Ambient Backscatter Communications to Protect Digital Content and Copyrights, IEEE Access 13 (2025) 27866–27877, https://doi.org/10.1109/ACCESS.2025.3632851.