Deep Reinforcement Learning for Adaptive Non-Primary Channel Access in IEEE 802.11bn

Taewon Song
Dept. of Internet of Things
Soonchunhyang University
Asan, Republic of Korea
twsong@sch.ac.kr

***Abstract*—Efficient spectrum utilization is critical in modern Wi-Fi networks as traditional systems require primary channel occupancy for transmission, limiting efficiency in overlapping BSS (OBSS) environments. IEEE 802.11bn introduces non-primary channel access (NPCA) capability, yet optimal decision strategies remain challenging. This paper presents a deep reinforcement learning approach for adaptive NPCA decision-making using Semi-Markov Decision Process formulation with Deep Q-Network. Simulations across varying network scenarios demonstrate significant throughput improvements over baseline strategies, with contention window index as the most critical decision factor. The learning algorithm exhibits conservative strategies favoring long-term stability, providing insights for next-generation Wi-Fi channel access mechanisms.**

***Index Terms*—Deep Reinforcement Learning, Non-Primary Channel Access, Wi-Fi Networks, Semi-MDP, OBSS, Channel Access, DQN**

I. Introduction

Modern wireless networks face increasing challenges in spectrum efficiency as Wi-Fi deployments become denser and user demands grow. Traditional channel access mechanisms, while effective in simple scenarios, struggle to adapt to dynamic interference patterns and varying network conditions.

IEEE 802.11 systems traditionally require the primary channel to be idle before wide-band transmissions can occur [1]. This constraint leads to significant spectrum waste when secondary channels remain unused despite primary channel occupancy by overlapping BSS (OBSS) traffic. While IEEE 802.11bn introduces non-primary channel access (NPCA) capability [2], existing approaches rely on static heuristics that cannot adapt to dynamic network conditions, leaving a critical gap in intelligent decision-making strategies.

Consider a scenario where a station detects OBSS activity on its primary channel while secondary channels are available. The station must decide whether to wait for primary channel access or switch to NPCA, balancing factors such as transmission duration, channel switching overhead, and future network conditions. Such decisions require adaptive intelligence beyond static rules.

In this paper, we describe an intelligent NPCA decision-making framework that enables stations to learn optimal channel access policies through interaction with dynamic network environments. We formulate this as an online learning problem where stations adapt their behavior based on observed network states and reward feedback.

Our approach employs deep reinforcement learning, specifically a Semi-Markov Decision Process (Semi-MDP) formulation with Deep Q-Network (DQN) [3], to capture temporal dependencies in NPCA decisions. The framework enables stations to learn from experience and adapt to varying OBSS patterns and network densities.

The main contributions of this work are:

- A Semi-MDP framework for NPCA decision-making that captures temporal dynamics and network state transitions
- A DQN-based learning algorithm that enables adaptive channel access policies in dynamic environments
- Comprehensive performance evaluation demonstrating throughput improvements over baseline strategies
- Analysis of key decision factors revealing the critical role of contention window index in NPCA decisions

The remainder of this paper is organized as follows. Section II reviews related work in NPCA and reinforcement learning applications. Section III presents our system model and problem formulation. Section IV describes the proposed DRL framework. Section V presents simulation results and analysis. Finally, Section VI concludes the paper and discusses future work.

II. Related Work

NPCA mechanisms have been extensively studied in the context of spectrum efficiency improvement. Traditional approaches rely on heuristic rules and static thresholds for channel switching decisions [4]. However, these methods fail to adapt to dynamic network conditions and varying traffic patterns.

Reinforcement learning has shown promising results in wireless network optimization [5]. Recent works have applied DRL to various wireless problems, including resource allocation and interference management. Semi-MDP formulations have been particularly effective in capturing temporal dependencies in wireless environments [6].

Existing NPCA studies focus primarily on theoretical analysis and static optimization. This work addresses the gap by proposing an adaptive learning approach that can respond to real-time network dynamics.

III. System Model and Problem Formulation

A. NPCA System Model

We consider a wireless network with overlapping BSSs where stations operate on a primary channel (Channel 0) and can access a secondary NPCA channel (Channel 1) when OBSS activity is detected. Each station maintains awareness of primary channel occupancy and makes binary decisions: stay on primary (Action 0) or switch to NPCA (Action 1).

B. Semi-MDP Formulation

We formulate the NPCA decision problem as a Semi-MDP with the following components:

State Space: The state s_t includes:

- Primary channel OBSS occupation time
- Radio transition time
- Transmission duration
- Contention window index

Action Space: Binary actions $a_t \in \{0, 1\}$ where 0 represents staying on primary channel and 1 represents switching to NPCA channel.

Reward Function: Rewards are based on successful PPDU transmission length, with zero reward for transmission failures.

The state transition probability follows:

$$P(s_{t+1}|s_t, a_t) = \mathbb{P}[\text{next state}|\text{current state, action}] \quad (1)$$

The Q-function update in our DQN framework is:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)] \quad (2)$$

IV. Proposed DRL Framework

A. Semi-MDP Learner Architecture

Our `SemiMDPLearner` class implements a DQN-based learning algorithm with experience replay and target network stabilization. The neural network architecture consists of fully connected layers with ReLU activations, mapping state observations to Q-values for each action.

B. Training Process

The learning process follows these key steps:

1) State observation and action selection using ϵ-greedy policy
2) Experience storage in replay memory buffer
3) Batch sampling and Q-network updates
4) Target network periodic synchronization
5) Exploration rate decay over episodes

C. Network Architecture and Hyperparameters

The DQN consists of three hidden layers (128, 128, 64 neurons) with dropout regularization. Key hyperparameters include learning rate $\alpha = 0.001$, discount factor $\gamma = 0.99$, and replay memory capacity of 10,000 transitions.

V. Simulation Results

A. Experimental Setup

Simulations are conducted using a time-slotted framework with slot duration of 9 μs following IEEE 802.11ax standards. We evaluate networks with varying numbers of stations per channel, comparing DRL-based NPCA against baseline approaches including offload-only and local-only strategies.

B. Performance Metrics

We measure:

- Throughput: Successful data transmission rate
- Channel utilization: Ratio of successful channel occupation
- Fairness: Inter-BSS performance balance
- Learning convergence: Episode reward progression

C. Performance Comparison

Results demonstrate that the DRL-based approach achieves superior performance compared to static strategies. Across different network scenarios, our method shows significant throughput improvement over baseline approaches including offload-only and local-only strategies. The learning algorithm effectively adapts to varying OBSS patterns and channel conditions.

D. Decision Factor Analysis

Analysis reveals that contention window index serves as the most critical decision factor, followed by OBSS occupation time. The learned policy exhibits conservative behavior, favoring long-term stability over aggressive short-term gains.

VI. Conclusion and Future Work

This paper presented a DRL-based approach for adaptive NPCA decision-making in IEEE 802.11bn networks. The Semi-MDP formulation with DQN learning enables stations to intelligently choose between primary and secondary channel access based on dynamic network conditions.

Key findings include the importance of contention window index as a decision factor and the effectiveness of conservative learning strategies. Future work will explore multi-agent learning scenarios and adaptive frame duration optimization based on real-time network conditions.

TABLE I
Simulation Parameters

Parameter	Value
Simulation time	500,000 μs
Slot duration	9 μs
Number of channels	2 $\times$ 20 MHz
STAs per channel	2, 6, or 10
Frame duration (Short)	33 slots (297 μs)
Frame duration (Long)	165 slots (1485 μs)
OBSS generation rate	0.05 per slot
NPCA switching delay	5 slots
Learning rate	0.001
Discount factor	0.99

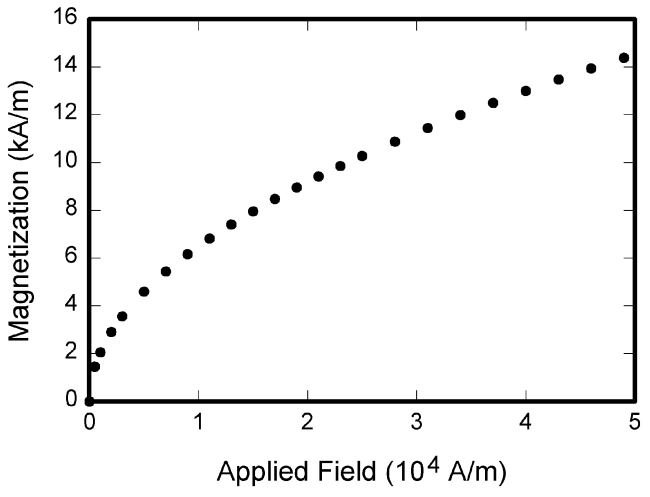

Fig. 1. Training convergence showing episode rewards over time for DRL-based NPCA learning in different network densities.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) in 2025 (2021-0-01399).

REFERENCES

[1] D. Wei, L. Cao, L. Zhang, X. Gao, and H. Yin, "Non-Primary Channel Access in IEEE 802.11 UHR: Comprehensive Analysis and Evaluation," in *2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall)*. IEEE, 2024, pp. 1–6.

[2] B. Bellalta, F. Wilhelmi, L. Galati-Giordano, and G. Geraci, "Performance Analysis of IEEE 802.11 bn Non-Primary Channel Access," *arXiv preprint arXiv:2504.15774*, 2025.

[3] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-Level Control Through Deep Reinforcement Learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[4] D. Wei, L. Cao, L. Zhang, X. Gao, and H. Yin, "Optimized Non-Primary Channel Access Design in IEEE 802.11 bn," in *GLOBECOM 2024-2024 IEEE Global Communications Conference*. IEEE, 2024, pp. 4588–4593.

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," *arXiv preprint arXiv:1312.5602*, 2013.

[6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, 2018.