

Title	Attention-enhanced DQN scheduling for multi-link devices in synchronous N-STR Wi-Fi 7 networks
Authors	Abdelreheem, Ahmed;Sreenan, Cormac;Zahran, Ahmed H.
Publication date	2025-12-30
Original Citation	Abdelreheem, A, Sreenan, C & Zahran, A H 2025, Attention-enhanced DQN scheduling for multi-link devices in synchronous N-STR Wi-Fi 7 networks. in 2025 International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM). MSWiM 2025 - 27th International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Institute of Electrical and Electronics Engineers Inc., pp. 659-666, 27th International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM 2025, Barcelona, Spain, 27/10/25. https://doi.org/10.1109/MSWiM67937.2025.11309124
Type of publication	Conference item
Link to publisher's version	https://www.scopus.com/pages/publications/105032535437 - 10.1109/MSWiM67937.2025.11309124
Rights	© 2025, IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Download date	2026-10-10 19:21:03
Item downloaded from	https://hdl.handle.net/10468/18804



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Attention-Enhanced DQN Scheduling for Multi-Link Devices in Synchronous N-STR Wi-Fi 7 Networks

Ahmed Abdelreheem^{1,2}

¹*University College Cork*
Cork, Ireland
aabdelreheem@ucc.ie

²*Electrical Engineering Department*
Faculty of Engineering, South Valley University
Qena, Egypt

Cormac Sreenan

School of computer Science and IT
University College Cork
Cork, Ireland
cormac.sreenan@ucc.ie

Ahmed H. Zahran

School of computer Science and IT
University College Cork
Cork, Ireland
ahmed.zahran@ucc.ie

Abstract—WiFi 7 IEEE (802.11be) introduced Multi-Link Operation (MLO) that enables its devices to communicate over multiple links to support evolving latency-sensitive and high-throughput applications. However, MLO requires advanced scheduling algorithms to optimize the operation. This paper models WiFi 7 scheduling as a constrained Markov Decision Process that optimizes throughput, delay, latency and fairness while capturing access constraints and traffic dynamics. We also develop an attention-enhanced Rainbow Deep Q-Network (DQN) scheduling framework that combines multi-head attention, distributional Q-learning, and prioritized experience replay. Simulation results show up to $2.3\times$ throughput improvement, $6\times$ delay reduction, and marked gains in packet drop rate and spectral efficiency over baseline Round Robin scheduling.

Index Terms—Wi-Fi 7, MLO, N-STR, DQN, DRL.

I. INTRODUCTION

IEEE 802.11be introduced significant enhancements to the physical (PHY) and medium access control (MAC) layers [1]. 4096-QAM modulation, 320 MHz bandwidth channels, up to 16 spatial streams, and access to the newly allocated 6 GHz band alongside legacy 2.4 and 5 GHz bands are unique features of WiFi 7 [2]. Additionally, Multi-Link Operation (MLO) stands out as another unique capability of Wi-Fi 7. In legacy Wi-Fi networks, devices operated in single-link mode, meaning they could transmit or receive over only one frequency band at a time, such as 2.4 GHz, 5 GHz, or 6 GHz. This operation limited throughput and made devices more vulnerable to congestion or interference on that single band.

MLO enables WiFi 7 Multi-Link Devices (MLDs) to establish and manage multiple simultaneous links across different frequency bands. By aggregating or dynamically switching between links, MLO improves throughput, latency, and reliability, and unlocks new capabilities like link steering, load balancing, and failover. These improvements enable high-throughput, low-latency requirements of evolving applications,

such as augmented/virtual reality (AR/VR), cloud gaming, and robotics [3]. However, the development of advanced scheduling algorithms is crucial to reaping the benefits of these new capabilities [4].

MLO supports different data communication modes based on transmission coordination and activity across different bands. Considering transmission coordination, data exchange can be classified as synchronous and asynchronous. In synchronous MLO, all active links are tightly coordinated to start and end transmission opportunities concurrently across all bands. Conversely, asynchronous MLO considers independent channel access and transmission timelines for all links. While asynchronous operation is more flexible, it requires complex hardware design. Considering transmission activity, WiFi 7 defines simultaneous transmit and receive (STR) and non-simultaneous transmit and Receive (N-STR) [5]. STR implies the devices capability to transmit and receive concurrently across multiple links, requiring advanced physical and RF design. On the contrary, N-STR is considered a practical and cost-effective configuration for Wi-Fi 7 deployment, where links are used in a synchronized manner either for transmission or reception, but not both simultaneously.

Synchronous N-STR mode reduces hardware complexity, lowers power consumption, and simplifies RF front-end design. However, it also may limit transmission opportunities if one link is busy due to contention or interference. This constraint imposes a stringent coordination requirement: all scheduled links must be free to avoid half-duplex conflicts or collisions. Such constrain may increase latency and reduce spectrum utilization. Hence, advanced multi-band scheduling is crucial to reaping the benefits of WiFi 7 in dense or dynamic environments [6].

When an access point (AP) acquires the medium, it typically schedules multiple stations (STAs) across the different bands. However, conventional scheduling strategies, such as Round Robin or static prioritization generally ignore temporal link states and inter-link coordination, leading to increased

delay, buffer overflows, and inefficient spectrum utilization [5], [7]–[10]. Optimized scheduling decisions are challenged by combinatorial explosion in the state space with the number of bands and users due to traffic and channel diversity [11], [12]. Addressing these challenges is central to unlocking the full potential of Wi-Fi 7 MLO in real-world deployments.

In this paper, we present a novel Deep Q-Network (DQN)-based scheduler for Wi-Fi 7 SN-STR. Our main contributions are as follows:

- We propose a multi-objective optimization framework for the scheduler in the N-STR system, designed to jointly optimize throughput, delay, packet drop, and fairness. The framework also ensures flexible configurability, enabling adaptation to diverse service requirements.
- We develop an attention-enhanced Rainbow DQN architecture that incorporates advanced reinforcement learning techniques, dueling networks, distributional Q-learning, prioritized experience replay, and noisy exploration for robust policy learning. To optimize performance trade-offs, we apply Bayesian optimization for principled reward weight tuning during training.
- We evaluate the proposed scheme using a high-fidelity Wi-Fi 7 simulation environment and benchmark it against conventional baseline methods. Simulation results demonstrate significant improvements in throughput, latency, and packet loss, thereby validating the effectiveness of DQN for next-generation wireless resource management in MLO scenarios.

This study offers new insights into intelligent scheduling for Wi-Fi 7 SN-STR and lays the groundwork for practical deployments of AI-driven MAC protocols in future wireless networks.

II. BACKGROUND AND RELATED WORK

A. N-STR Operation

In legacy Wi-Fi networks, devices compete for the medium access using carrier sense multiple access with collision avoidance (CSMA/CA) that involves sensing and random backoff elements. Wi-Fi 7 extended this single-link mechanism to enable MLO. Two sensing frameworks are proposed, including single primary channel (SPC) and multiple primary channel (MPC). SPC contends for the access on a selected primary band and briefly checks the other bands if a transmission opportunity is deemed possible. Alternatively, MPC scans all bands to ensure that they are idle. MPC may use the shortest or longest backoff timer when contending on multiple bands. MPC is more complex to implement but offers better performance in congested environments.

Once a transmission opportunity is acquired, the MLD schedules the packets to be transmitted in every band. To ensure synchronized MLO, padding is used to align the transmission durations over multi-link by appending extra bits to shorter frames. Padding can lead to airtime waste and could be critical for heterogeneous traffic. In the uplink, MLD AP leverage Trigger Frames (TFs), extended in IEEE 802.11be, to

probe buffer occupancy at STAs. The AP then schedules STAs based on its policy. In the presence of legacy devices, they may contend to access the medium. Therefore, a combination of CSMA/CA-aware contention management, uplink scheduling, and adaptive padding control is essential to enable reliable, high-performance MLO in Wi-Fi 7 networks.

N-STR MLDs offer distinct advantages, including lower power consumption, simplified radio architecture, and reduced manufacturing costs, attributes especially beneficial for commercial devices. Although incapable of simultaneous multi-link operation, N-STR MLDs still deliver key MLO benefits such as reduced latency, improved reliability through link diversity, and enhanced channel utilization over legacy single-link designs. Their streamlined hardware simplifies RF shielding and mitigates electromagnetic interference (EMI). These combined factors make N-STR a compelling option. Consequently, while N-STR provides tangible implementation advantages, it entails a trade-off in scheduling flexibility and efficiency. The IEEE 802.11be draft leaves scheduling policies open, allowing vendors and researchers significant latitude in developing optimization strategies for traffic distribution across multiple links.

B. MLO Scheduling

Recent advancements in IEEE 802.11be (Wi-Fi 7) have led to significant exploration of MLO and scheduling mechanisms. Lopez-Ravents and Bellalta [2] proposed a dynamic traffic allocation framework that shows adaptive link assignment significantly outperforms static methods in terms of congestion control and delivery fairness. Complementarily, [13] proposed a scalable MAC with AP coordination and full-duplex multi-band support, achieving notable throughput gains by leveraging aggregation and bonding. The service level agreements (SLAs)-MLO framework [11] introduced a congestion-aware scheduler that dynamically maps flows to links by SLA, reducing SLA deviations by up to 50% under various congestion levels. For latency-sensitive traffic, Bankov et al. [12] showed that isolating high-priority flows on dedicated links cuts delay while maintaining best-effort performance. Further work by Lopez-Ravents and Bellalta [5] found that congestion-aware single-link allocation outperforms naive splitting in dense WLANs. An analytical framework proposed in [6] explored how traffic allocation, idle listening, and channel diversity affect flow completion times.

Machine learning has also emerged as a design methodology for MLO. Dawar et al. [7] used traffic classification and QoE-aware scheduling to optimize link utilization and energy efficiency. Lv et al. [8] analyzed blocking and idle dynamics for IEEE 802.11be MLO. Other studies [9], [10], [14]–[16] focused on mitigating interference with legacy devices and improving overall system stability through proposed coexistence-aware access with dynamic contention tuning.

Reinforcement learning (RL) has driven innovations such as deep RL-based frame aggregation [17], multi-agent RL for STR-based access [18], and centralized DRL for throughput-aware scheduling [19]. A hybrid Multi-Armed Bandit (MAB)

Synchronous Multi-Link Operation

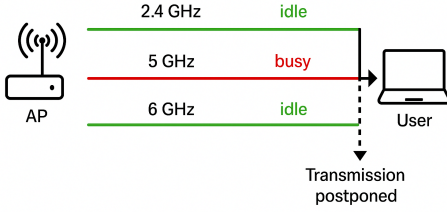


Fig. 1: Non-Simultaneous Transmit and Receive (N-STR) MLD.

learning with Large Language Models (LLMs) framework [20] further supports intelligent channel allocation in dynamic WLANs. However, most aforementioned work consider an individual objective or adopted a static methodology. In contrast, our work considers a multi-objective strategy that facilitates adaptive QoS-aware scheduling.

III. SYSTEM MODEL

A. Network Model

We consider a Wi-Fi 7 (IEEE 802.11be) network comprised of an AP operating in N-STR MLD mode over n independent wireless bands, indexed by $\mathcal{B} = \{1, 2, \dots, n\}$, e.g., 2.4 GHz, 5 GHz, and 6 GHz. Each band $b \in \mathcal{B}$ serves a set $\mathcal{U}_b$ of $|\mathcal{U}_b|$ STAs. The AP maintains finite-capacity buffers for each STA with Poisson packet arrivals at rate $\lambda_{b,u}$ packets per slot. Following the N-STR operating principles, a transmission opportunity is only initiated when all bands are simultaneously idle; otherwise, no transmission occurs. We consider a time-slotted system where a time slot corresponds to a transmission opportunity. Hence, the AP acquires a transmission opportunity when

$$I_b[t] = 1, \forall b \quad (1)$$

where $I_b[t]$ indicates whether band b is sensed idle at slot t .

Let $x_{b,u}[t] = 1 \in \{0, 1\}$ indicate that STA u on band b is scheduled at slot t . When the AP acquires the transmission opportunity for slot t , its scheduler decides one or more users to serve in each band. Without loss of generality, we assume for each band b and time t , at most one user is scheduled. This assumption avoids intra-band interference, maximizes per-user transmission rates through full bandwidth allocation, reduces scheduling complexity from exponential to linear in the number of users, and focuses optimization on the core N-STR challenge of cross-band coordination. Thus, the joint scheduling action is:

$$x_{b,u}[t] = \begin{cases} 1, & \text{if } I_b[t] = 1, \\ 0, & \text{otherwise} \end{cases} \quad \forall b, t \quad (2)$$

$$\sum_{u \in \mathcal{U}_b} x_{b,u}[t] \leq 1, \quad \forall b, t \quad (3)$$

The instantaneous signal-to-noise ratio (SNR) for user u on band b at slot t , $SNR_{b,u}[t]$, evolves as a temporally correlated Markov process [21]:

$$SNR_{b,u}[t] = \rho \cdot SNR_{b,u}[t-1] + (1-\rho) [\mu_{b,u}[t] + \sigma Z_{b,u}[t]] \quad (4)$$

where ρ is the channel correlation coefficient, $\mu_{b,u}$ is the time-varying mean SNR, σ is the fading standard deviation, and $Z_{b,u}[t] \sim \mathcal{N}(0, 1)$ models i.i.d. Gaussian noise.

In this work, all wireless impairments, including interference, attenuation, and noise, are abstracted into the SNR variable, which evolves as a temporally correlated random process per user and band. This approach provides a tractable and commonly used abstraction for a single AP Wi-Fi 7 deployment. For more complex topologies such as dense networks or multi-AP scenarios, our framework can be extended to incorporate explicit path loss, shadowing, and inter-cell interference models by refining the SNR process accordingly. This is left for future work.

The Modulation and Coding Scheme (MCS) for user u in band b at time slot t , denoted $m_{b,u}[t]$, is mapped from SNR per the IEEE 802.11be specification [1]. In this work, all wireless impairments, including interference, attenuation, and noise, are abstracted into the SNR variable. This approach provides a tractable and commonly used abstraction for single AP Wi-Fi 7 deployments.

B. Buffer Dynamics

The AP maintains a dedicated downlink buffer for each STA. The buffer dynamics are governed by random arrival and transmission processes in every time slot. Let $\lambda_{b,u}[t]$ and $Thr_{b,u}[t]$, respectively, represent the number of packets arrived and transmitted for user u on band b in time slot t . The user buffer size, denoted as $Q_{b,u}[t]$, is updated as:

$$Q_{b,u}[t+1] = \max\{0, Q_{b,u}[t] - Thr_{b,u}[t]\} + \lambda_{b,u}[t] \quad (5)$$

$Thr_{b,u}[t]$ can be expressed as

$$Thr_{b,u}[t] = \left\lfloor \frac{N_{DBPS}(m_{b,u}[t]) \cdot N_{sym}}{L_{packet} \cdot 8} \right\rfloor \quad (6)$$

where $N_{DBPS}(m)$ is the number of data bits per OFDM symbol for MCS level m :

$$N_{DBPS}(m) = b_{pscs}(m) \cdot r(m) \cdot RU_{b,u}[t] \quad (7)$$

$b_{pscs}(m)$ is the bits per subcarrier symbol for MCS m (e.g., 1 for BPSK, 2 for QPSK), $r(m)$ is the coding rate for MCS m (e.g., 1/2, 3/4, 5/6). $RU_{b,u}[t]$ is the number of tones in the assigned resource unit. $N_{sym} = 73$ is the number of OFDM symbols per time slot (1 ms with 13.6 μ s symbol duration). $L_{packet} = 1500$ bytes is the packet size.

Considering the limited buffer capacity, denoted as $Q_{b,u}^{max}$, the number of dropped packets for user u in band b during time slot t , denoted as $Drop_{b,u}[t]$, can be expressed as:

$$Drop_{b,u}[t] = \max\{0, Q_{b,u}[t] + \lambda_{b,u}[t] - Q_{b,u}^{max}\} \quad (8)$$

IV. MLO SCHEDULE OPTIMIZATION

Our N-STR scheduler optimizes a multi-objective utility that rewards long-term throughput and fairness and penalizes delay and packet drops. For fairness, we consider Jains Fairness Index, expressed as:

$$F_{Jain}[t] = \frac{\left(\sum_{b,u} Thr_{b,u}[t]\right)^2}{BU \sum_{b,u} Thr_{b,u}^2[t]} \quad (9)$$

The value of Jains fairness index ranges from $1/(BU)$ (highly unfair) to 1 (perfect fairness). The delay component is more relevant to time-sensitive applications that operate on strict delay budgets. Let $D_{b,u}[t]$ is the head-of-line (HoL) packet buffering delay for user u in band b at tile t . The delay constraint for these applications is defined as:

$$D_{b,u}[t] \leq L_{b,u}, \quad \forall t \quad (10)$$

where $L_{b,u}$ represents the maximum delay bound.

At each N-STR transmission opportunity, the scheduler selects a user in each band to maximize the overall utility across a finite scheduling horizon (T slots). The scheduling optimization problem is defined as follows.

$$\max_{x_{b,u}[t]} \sum_{t=1}^T \left[\sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}_b} \alpha_{b,u} Thr_{b,u}[t] + \beta F_{Jain}[t] - \gamma \frac{D_{b,u}[t]}{L_{b,u}} x_{b,u}[t] - \delta Drop_{b,u}[t] \right] \quad (11)$$

Subject to:

$$\sum_{u \in \mathcal{U}_b} x_{b,u}[t] \leq I_b[t], \quad \forall b, t \quad (C1)$$

$$RU_{b,u}[t] = R_b x_{b,u}[t] \quad (C2)$$

$$Q_{b,u}[t+1] = \max\{0, Q_{b,u}[t] - Thr_{b,u}[t]\} + \lambda_{b,u}[t] \quad (C3)$$

$$Drop_{b,u}[t] = \max\{0, Q_{b,u}[t] + \lambda_{b,u}[t] - Q_{b,u}^{max}\} \quad (C4)$$

$$D_{b,u}[t] \leq L_{b,u}, \quad \forall t \quad (C5)$$

$$x_{b,u}[t] \in \{0, 1\}, \quad \forall b, u, t \quad (C6)$$

$\alpha_{b,u}$, β , γ , and δ are weights for objective components.

The scheduling problem can be modeled as a constrained Markov Decision Process (CMDP) [22]. The environment state $S[t]$ includes, for all users and bands, the current queue lengths, SNR values, head-of-line delays, historical buffer statistics, recent channel quality, drop rates, and global system features, such as average queue and SNR per band. Action $\mathcal{A}[t]$: at each transmission opportunity, the agent selects one STA per band to schedule for transmission; this is represented as a vector of binary scheduling variables $x_{b,u}[t]$. Reward $r[t]$: the reward function captures throughput, fairness, delay penalty, and packet drops, guiding the agent to maximize the overall scheduling utility. Transition: The environment updates the state according to buffer dynamics, channel evolution (including correlated SNR variations), and packet arrivals, as previously detailed. Hence, advanced DRL approaches can be leveraged to optimize N-STR scheduling in Wi-Fi 7 networks.

V. PROPOSED SCHEME: DEEP REINFORCEMENT LEARNING-BASED SCHEDULING

To solve the N-STR scheduling problem, we propose a DRL framework that based on the Rainbow DQN architecture augmented with an attention mechanism [23]. This approach enables the AP to learn optimal scheduling decisions in a highly dynamic Wi-Fi 7 multi-band environment with diverse QoS demands, stochastic traffic arrivals, and temporally correlated fading channels.

A. Rainbow DQN with Attention Architecture

To address the high-dimensional state and action space in multi-band, multi-user Wi-Fi 7 scheduling, we adopt a deep reinforcement learning framework based on the Rainbow DQN algorithm, further enhanced by a multi-head attention mechanism. The proposed agent is designed to process complex cross-user and cross-band interactions, efficiently learn optimal scheduling strategies, and maximize the constrained utility function described in the previous sections.

At each discrete time step t , the environment presents the agent with a global system state $S[t]$, which concatenates per-user features including $Q_{b,u}[t]$, $SNR_{b,u}[t]$, $D_{b,u}[t]$, $m_{b,u}[t]$, as well as global features such as mean queue length and spectral efficiency across all users and bands. The state can be represented as a matrix $\mathbf{F}[t] \in \mathbb{R}^{B \times U \times d}$, where d is the feature dimension. Data collection operates within existing IEEE 802.11be MLO frameworks with no additional control plane overhead, and most state information is locally maintained at the AP. The networks input layer first projects these raw features into a latent representation using a fully connected mapping [24]:

$$\mathbf{H}^{(0)} = \varphi(\mathbf{F}[t] \mathbf{W}_0 + \mathbf{b}_0), \quad (12)$$

where $\mathbf{W}_0$ and $\mathbf{b}_0$ are learnable parameters, and $\varphi(\cdot)$ denotes a non-linear activation function. Subsequently, a multi-head self-attention block is applied to model the dependencies between users and bands. The attention mechanism computes, for each head $k \in \{1, \dots, K\}$ as follows [25]:

$$\text{Attention}^{(k)}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q} \mathbf{K}^\top}{\sqrt{d_k}} \right) \mathbf{V}, \quad (13)$$

where $\mathbf{Q} = \mathbf{H}^{(0)} \mathbf{W}_Q^{(k)}$, $\mathbf{K} = \mathbf{H}^{(0)} \mathbf{W}_K^{(k)}$, $\mathbf{V} = \mathbf{H}^{(0)} \mathbf{W}_V^{(k)}$, and d_k is the head dimension. The outputs from all heads are concatenated and linearly projected to form the attention-enhanced feature map [25]:

$$\mathbf{H}^{(1)} = \text{Concat}_{k=1}^K \left(\text{Attention}^{(k)}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \right) \mathbf{W}_0, \quad (14)$$

This attended feature representation is passed through additional fully connected layers, yielding a shared latent embedding $\mathbf{Z}$ that captures global network context and user-specific conditions.

The dueling DQN architecture is then employed to separately estimate the state-value function $V(S[t])$ and the advantage function $\mathcal{A}(S[t], \mathcal{A}[t])$ for each possible scheduling

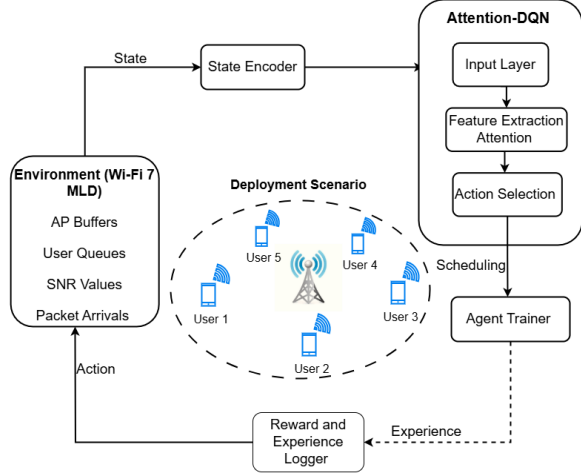


Fig. 2: Rainbow DQN with Attention scheduling framework for Wi-Fi 7 multi-band networks.

action $\mathcal{A}[t]$. Specifically, the final Q-value for action $\mathcal{A}[t]$ is given by [26]:

$$Q(S[t], \mathcal{A}[t]) = V(S[t]) + \left(\mathcal{A}(S[t], \mathcal{A}[t]) - \frac{1}{|\mathcal{A}|} \sum_{\mathcal{A}'} \mathcal{A}(S[t], \mathcal{A}') \right), \quad (15)$$

where $|\mathcal{A}|$ is the number of feasible scheduling actions. The network outputs a distribution over value atoms, $\{z_i\}_{i=1}^N$, representing the probability masses of potential return values, as required by the distributional RL component of Rainbow DQN. Letting $p_i(S[t], \mathcal{A}[t])$ denote the probability assigned to atom z_i , the expected Q-value is [27]:

$$Q(S[t], \mathcal{A}[t]) = \sum_{i=1}^N z_i p_i(S[t], \mathcal{A}[t]). \quad (16)$$

The training objective minimizes the cross-entropy between the predicted and projected target distributions, where the multi-step target for each transition $(S[t], \mathcal{A}[t], r[t], S[t+1])$ is computed as [27]:

$$T_z = r[t] + \omega^n \max_{\mathcal{A}'} Q(S[t+n], \mathcal{A}'), \quad (17)$$

with n being the n -step return parameter and ω the discount factor. The agent employs prioritized experience replay, sampling transitions with probability proportional to their temporal-difference error. For exploration, the noisy network layers inject trainable stochasticity into the Q-value estimates, facilitating efficient discovery of optimal policies. The target network parameters θ^- are updated periodically using a soft update rule [27]:

$$\theta^- \leftarrow \tau \theta + (1 - \tau) \theta^-, \quad (18)$$

where $\tau \in (0, 1)$ is the update rate. The target network soft update is performed every f_{target} training steps to maintain stability while allowing gradual parameter evolution. Training commences after an initial exploration period of $\varepsilon_{\text{init}}$ episodes,

Algorithm 1 Bayesian Optimized Rainbow DQN with Multi-Head Attention

- 1: **Input:** $\mathbf{F}[t]$ including environment parameters $(\mathcal{B}, \mathcal{U}_b, E, T, \text{SNR model})$
- 2: Initialize DQN θ , target θ^- , $\mathcal{D}$, $\mathcal{B}_N$
- 3: Initialize Bayesian optimizer $\mathcal{BO}$ $\mathbf{w} = [w_1, w_2, w_3, w_4]$
- 4: **for** episode $e = 1$ to E **do**
- 5: **for** slot $t = 1$ to T **do**
- 6: Update STA states, extract $S[t]$, get available $\mathcal{A}_t$
- 7: Select action: $a_t \leftarrow \arg \max_{a \in \mathcal{A}_t} Q(s_t, a; \theta)$
- 8: Execute a_t , calculate reward: $r_t = \sum_{i=1}^4 w_i$.
- 9: Store (s_t, a_t, r_t, s_{t+1}) in $\mathcal{B}_N$; if full, transfer to $\mathcal{D}$
- 10: **end for**
- 11: **if** $e > \varepsilon_{\text{init}}$ **then**
- 12: Train: sample from $\mathcal{D}$, update θ , soft update θ^- every f_{target} steps
- 13: **end if**
- 14: **if** $e \geq \varepsilon_0$ AND $e \bmod \Delta_{\text{BO}} = 0$ **then**
- 15: Update $\mathcal{BO}$, suggest $\mathbf{w}^{\text{new}}$ via Expected Improvement, normalize weights
- 16: **end if**
- 17: **end for**
- 18: **Output:** Optimal policy π^* and weights $\mathbf{w}^*$

ensuring sufficient experience accumulation in the replay buffer before gradient-based optimization begins.

Through this architecture, the Rainbow DQN with Attention agent learns to map the complex, temporally and spatially correlated state of the Wi-Fi 7 network to high-quality scheduling decisions, effectively maximizing the CMDP objective and ensuring robust, fair, and efficient use of multi-band wireless resources. The DRL workflow summary for the whole system is illustrated in the algorithm and as shown in Fig 2.

B. Training Methodology and Convergence Analysis

The training methodology and convergence behavior of the proposed attention-enhanced Rainbow DQN agent were evaluated through empirical analysis. Key performance indicators episodic rewards, throughput, delay, and fairness were tracked over 150 training episodes. Convergence was assessed using a 25 episode moving average to distinguish sustained learning trends from transient fluctuations. Quantitative metrics included improvements in average throughput between initial and final quartiles and performance stability measured via the standard deviation of throughput over the last 20% of episodes. The progression of cumulative rewards served as the primary indicator of policy refinement throughout the training process.

The implementation integrates prioritized experience replay with importance sampling, where transition probabilities are scaled by temporal-difference errors to enhance sample efficiency. Importance sampling weights are annealed from $\beta_{\text{DQN}} = 0.4$ to 1.0 over 100 episodes to reduce gradient bias while preserving effective exploration. For exploration, the agent employs parametric noise via NoisyLinear layers using a factorized Gaussian noise model. This state-dependent

TABLE I: Evaluation Setup

Parameter	Description/Value
RUs	996 tones (full-band per user, OFDMA)
t (time slot)	1ms
$Q_{b,u}[t]$	50 packets
$\lambda_{b,u}[t]$	Poisson (0.5 packets/slot)
$L_{b,u}$	10ms
Number of Episodes E	150
Number of Slots per Episode T	300
ε_0	30
Δ_{BO}	25
N_{eval}	20
Replay Buffer $\mathcal{D}$	100,000 transitions
$\mathcal{B}_N$ capacity	3 transitions
Number of Attention Heads K	4
Attention Head Dimension d_k	16
Hidden Layers	2–3 FC (64–128)
f_{target}	500
ε_{init}	10
Activation Function	ReLU
Optimizer	Adam

approach eliminates the need for explicit ϵ -greedy scheduling, enabling adaptive exploration aligned with the current state distribution. Noise parameters are reset after each gradient update to maintain exploration diversity throughout training.

The proposed system leverages Bayesian Optimization to automate the tuning of the objective reward function in wireless scheduling, $\mathbf{w}^* = [\alpha, \beta, \gamma, \delta]$ with simplex constraint, $\sum \mathbf{w}^* = 1$. It uses a Gaussian Process to model performance, requiring N_{eval} simulation episodes per weight configuration for reliable evaluation [28]. Starting from uniform weights during the first ε_0 episodes, the algorithm then iteratively refines weight configurations via Expected Improvement every Δ_{BO} episodes, constrained within practical bounds where throughput weights range as $\alpha \in [0.10, 0.70]$, $\beta \in [0.10, 0.60]$, $\gamma \in [0.10, 0.60]$, and $\delta \in [0.05, 0.40]$. The optimization problem is formulated as $\mathbf{w}^* = \arg \max f(\mathbf{w})$, where $f(\mathbf{w})$ represents the expected performance of the scheduling system under weight configuration. The Bayesian Optimization initialized multi-start local search with five restarts, ensuring adherence to simplex constraints while systematically balancing exploration and exploitation through fifty candidate evaluations per iteration. The method automatically discovers effective weight configurations without manual tuning, offering robust performance, though specific convergence times depend on the complexity of the performance landscape.

These results confirm the effectiveness and reliability of the attention-enhanced Rainbow DQN architecture for real-world Wi-Fi 7 scheduling scenarios.

VI. PERFORMANCE EVALUATION

A. Evaluation Setup

We implement the proposed DQN-based N-STR scheduler in PyTorch and simulate a Wi-Fi 7 MLD network across the 2.4, 5, and 6 GHz bands. Each band supports up to 20 stations with Poisson-modeled traffic arrivals ($\lambda_{b,u} = 0.5$ packets/slot), temporally correlated SNR fading where the initial mean SNR is drawn from $\mu_{b,u}[0] \sim \mathcal{U}(4, 43)$ dB and evolves over time

with correlation coefficient $\rho = 0.9$ and fading standard deviation $\sigma = 2$ dB. The system and DQN hyper-parameters are listed in Table I.

The proposed approach is benchmarked against the standard Round Robin scheduler. Key evaluation metrics include aggregate throughput, average packet delay, Jain's fairness index, and packet drop rate.

B. Results and Discussion

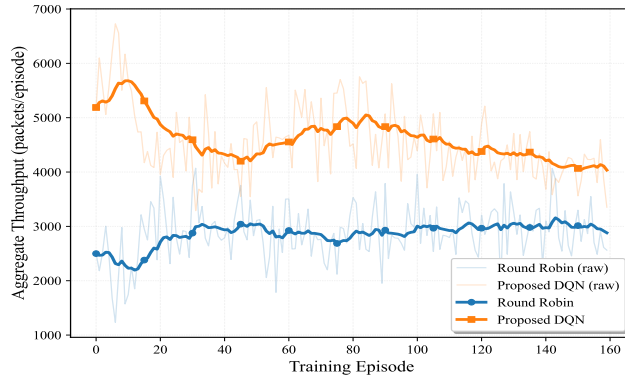
Fig. 3 plots both raw training measurements (transparent) and moving average trends (solid lines) calculated using a 25-episode sliding window. The DQN-based scheduler demonstrates clear superiority over the Round Robin baseline across all metrics. The DQN agent boosts the throughput by up to a $2.3\times$. The DQN scheduler consistently serves more packets per episode than RR (~ 6700 vs ~ 2900 packets/episode). This gain is observed to be stable as the network traffic load and environmental conditions vary, as shown in Fig.3a.

Additionally, Fig. 3b shows our DQN agent reduces the average packet delay to ~ 13 time slots in comparison to ~ 90 time slots for the Round Robin scheduler. This $6\times$ reduction enables meeting the stringent QoS requirements of latency-sensitive flows, which satisfies the AR/VR QoS requirement of < 20 ms end-to-end latency.

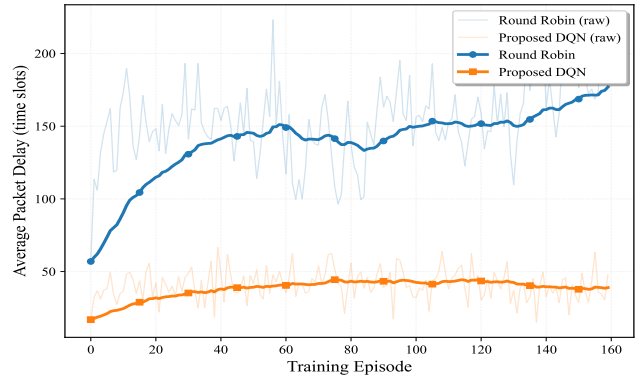
Fig. 3d illustrates that our DQN scheduler reduces packet drop rates to 3% , a huge reduction from the 25% achieved by the Round Robin scheduler. This reduction demonstrates the benefit of queue and channel-aware decision making to preserve the quality of service of the application layer.

Fig. 3c shows that our DQN scheduler converges to a fairness index of 0.6 in comparison to ~ 0.65 for the RR scheduler. It is important to note that the RR scheduler achieves a significantly lower throughput. More critically, this performance results from balancing the tradeoff between the different components of our scheduling objective. However, our scheduler behavior can be altered by tuning the reward function. Fig. 4 shows the Bayesian optimization process achieved convergence within five evaluation cycles, initially explored sub-optimal configurations, with performance decreasing from 0.66 to 0.57 at step 2, before discovering the optimal solution yielding a peak performance of 0.77 at step 4, representing a 16.7% improvement over uniform weighting. The discovered optimal weight configuration reveals delay minimization as the highest priority (50%), packet drop (20%), throughput (15%), and fairness (15%). This distribution suggests that delay minimization serves as an effective proxy for overall system optimization. The weight evolution trajectories demonstrate systematic learning, with throughput weights decreasing from initial exploration peaks while delay weights increase substantially to reach optimal values. The fairness and packet drop components exhibit more stable evolution, indicating robust constraint satisfaction throughout the optimization process.

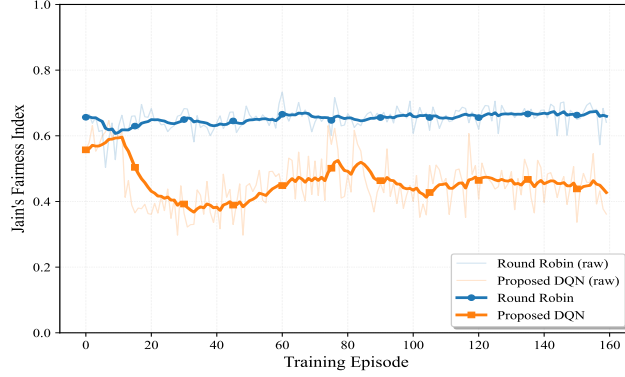
The attention mechanism significantly accelerates convergence and improves performance under correlated fading, demonstrating that the attention-enhanced DQN scheduler



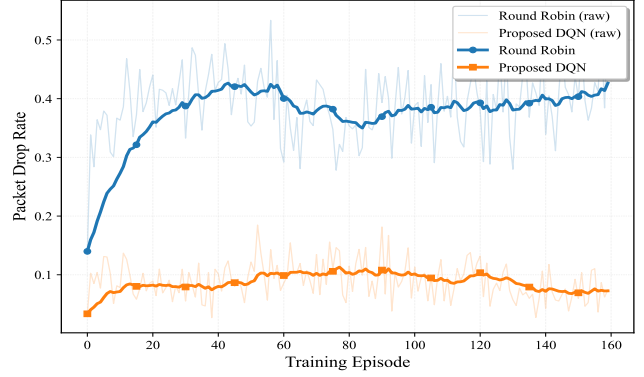
(a) Throughput



(b) Delay.



(c) Fairness.



(d) Packet Dropping.

Fig. 3: Performance metrics for tested scheduling Algorithms.

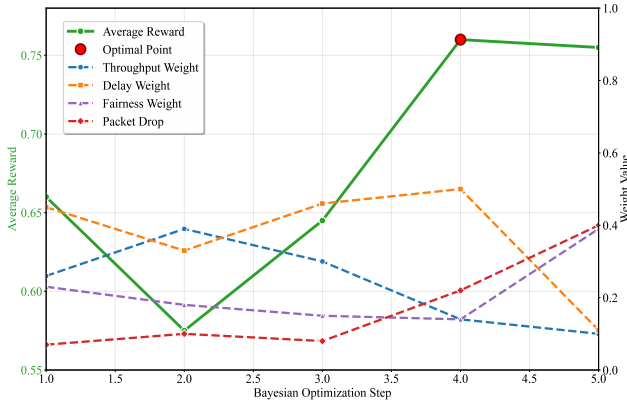


Fig. 4: Bayesian Optimization: Weight Evolution.

delivers superior multi-objective results in realistic Wi-Fi 7 N-STR environments, surpassing baseline methods in throughput, delay, fairness, and packet drop rate.

VII. CONCLUSION

This work presents a DRL framework for scheduling in Wi-Fi 7 synchronous N-STR MLD networks. Our proposed attention-enhanced Rainbow DQN scheduler considers a detailed network state while optimizing throughput, fairness, delay, and packet dropping. Simulation results demonstrate

up to $2.3\times$ throughput improvement, $6\times$ delay reduction, and sustained packet drop rates below 3%, in comparison to typical multi-user Round Robin baseline 25%. Additionally, Jain's fairness improve from 0.4 to 0.65 without sacrificing performance. The agent converges stably within 80 – 100 episodes. This work lays a foundation for intelligent, adaptive MAC-layer scheduling in Wi-Fi 7, addressing cross-band coordination in N-STR systems. Future research will explore uplink extensions and multi-AP deployments.

ACKNOWLEDGMENT

This research is supported by CONNECT—The Science Foundation Ireland (SFI) Research Centre for future networks (Grant: 13/RC/2077_P2).

REFERENCES

- [1] "IEEE Draft Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment: 320MHz Positioning," *IEEE P802.11bk/D4.0*, January 2025, pp. 1–116, 2025.
- [2] . Lpez-Ravents and B. Bellalta, "Dynamic Traffic Allocation in IEEE 802.11be Multi-Link WLANs," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1404–1408, 2022.
- [3] M. Alsakati, C. Pettersson, S. Max, V. N. Moothedath, and J. Gross, "Performance of 802.11be Wi-Fi 7 with Multi-Link Operation on AR Applications," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6.

- [4] E. Khorov, I. Levitsky, and I. F. Akyildiz, "Current Status and Directions of IEEE 802.11be, the Future Wi-Fi 7," *IEEE Access*, vol. 8, pp. 88 664–88 688, 2020.
- [5] J. Lpez-Ravents and B. Bellalta, "IEEE 802.11be Multi-Link Operation: When the Best Could Be to Use Only a Single Interface," in *2021 19th Mediterranean Communication and Computer Networking Conference (MedComNet)*, 2021, pp. 1–7.
- [6] M.-T. Suer, C. Thein, H. Tchouankem, and L. Wolf, "Multi-Connectivity as an Enabler for Reliable Low Latency Communications An Overview," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 156–169, 2020.
- [7] N. Dawar, K. N. Nguyen, A. Sehgal, Y. Zhu, B. Loong Ng, and J. Choi, "Enhancing Wi-Fi 7: Traffic Flow Intelligence and Multi-Link Operation for Optimal Efficiency," *IEEE Access*, vol. 13, pp. 63 298–63 309, 2025.
- [8] Y. Lv, Y. Chen, J. Li, M. Bi, and Y. Lu, "Multi-Link Operation With Exponential Increasing and Exponential Decreasing Hybrid Contention Windows for Low-Latency WLANs," *IEEE Communications Letters*, vol. 29, no. 4, pp. 834–838, 2025.
- [9] W. Murti and J.-H. Yun, "Multi-Link Operation with Enhanced Synchronous Channel Access in IEEE 802.11be Wireless LANs: Coexistence Issue and Solutions," *Sensors*, vol. 21, no. 23, p. 7974, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/23/7974>
- [10] —, "Multilink Operation in IEEE 802.11be Wireless LANs: Backoff Overflow Problem and Solutions," *Sensors*, vol. 22, no. 9, p. 3501, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/9/3501>
- [11] S. Kumar, E. Garcia-Villegas, and D. Camps-Mur, "SLA-MLO: Congestion-Aware SLA-Based Scheduling of Multiple Links in IEEE 802.11 be," in *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, 2024, pp. 875–880.
- [12] D. Bankov, A. Lyakhov, E. Khorov, and K. Chemrov, "On the Use of Multilink Access Methods to Support Real-Time Applications in Wi-Fi Networks," *Journal of Communications Technology and Electronics*, vol. 66, no. 12, pp. 1341–1350, 2021. [Online]. Available: <https://link.springer.com/article/10.1134/S1064226921120056>
- [13] M. Yang, B. Li, Z. Yan, and Y. Yan, "AP Coordination and Full-duplex enabled Multi-band Operation for the Next Generation WLAN: IEEE 802.11be (EHT)," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2019, pp. 1–7.
- [14] N. Korolev, I. Levitsky, I. Startsev, B. Bellalta, and E. Khorov, "Study of Multi-Link Channel Access Without Simultaneous Transmit and Receive in IEEE 802.11be Networks," *IEEE Access*, vol. 10, pp. 126 339–126 351, 2022.
- [15] L. Kwon and E.-C. Park, "Contention-Less Multi-Link Synchronous Transmission for Throughput Enhancement and Heterogeneous Fairness in Wi-Fi 7," *Sensors*, vol. 24, no. 11, p. 3642, 2024. [Online]. Available: <https://www.mdpi.com/1424-8220/24/11/3642>
- [16] J. Lpez-Ravents and B. Bellalta, "Multi-Link Operation in IEEE 802.11be WLANs," *IEEE Wireless Communications*, vol. 29, no. 4, pp. 94–100, 2022.
- [17] J. Wu, X. Fang, and G. Min, "Deep Reinforcement Learning Based Multi-Link Frame Aggregation Length Optimization in Next Generation Wi-Fi Networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 14 482–14 497, 2024.
- [18] B. Tan, Y. Gao, and X. Sun, "Throughput-Optimal Multi-Link Access for Wi-Fi 7 via Multi-Agent Reinforcement Learning," in *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, 2025, pp. 1–6.
- [19] L. Zhang, H. Yin, S. Roy, L. Cao, X. Gao, and V. Sathya, "IEEE 802.11be Network Throughput Optimization With Multilink Operation and AP Controller," *IEEE Internet of Things Journal*, vol. 11, no. 13, pp. 23 850–23 861, 2024.
- [20] S. Lian, J. Tong, J. Zhang, and L. Fu, "Intelligent Channel Allocation for IEEE 802.11be Multi-Link Operation: When MAB Meets LLM," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2025.
- [21] R. Játiva and J. Vidal, "Cramér-rao bounds in the estimation of time of arrival in fading channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 1, pp. 1–13, 2018. [Online]. Available: <https://link.springer.com/article/10.1186/s13634-018-0540-1>
- [22] L. N. Alegre, A. L. Bazzan, and B. C. d. Silva, "Reinforcement Learning for Multi-Objective and Constrained Markov Decision Processes," 2019. [Online]. Available: <https://arxiv.org/abs/1901.08978>
- [23] X. Li *et al.*, "An overview: Attention mechanisms in multi-agent reinforcement learning," *Neurocomputing*, vol. 611, p. 128860, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0925231224007860>
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, pp. 5998–6008, 2017.
- [25] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer," *International conference on machine learning*, pp. 4055–4064, 2018.
- [26] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," pp. 1995–2003, 2016.
- [27] P. S. Castro, S. Moitra, C. Gelada, S. Kumar, and M. G. Bellemare, "Dopamine: A research framework for deep reinforcement learning," 2018, open-source implementation of Rainbow DQN with equations 19–21.
- [28] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016.