JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-20, No.-8, August (2025) pp 11 - 35 ISSN (Print) 0973-8975

A DEEP REINFORCEMENT LEARNING APPROACH TO JOINT CODEBOOK SELECTION AND UE SCHEDULING FOR NR-U/WIGIG COEXISTENCE IN UNLICENSED MMWAVE BANDS

K N S K Santhosh[1] , Angara Satyam[2], Kante Satyanarayana[3], Venkata Raju Athili[4], Ponugoti Gangadhara Rao[5], Bhatraju Mahalakshmi Rao[6]

[1]Department of Computer Science and Engineering, Aditya University, Surampalem, India.
[2]Department of Computer Science and Design, SRKR Engineering College, Bhimavaram, India.
[3]Department of Computer Science and Engineering, Aditya University, Surampalem, India.
[4]Department of Computer Science and Engineering, International School of Technology and Sciences for Women, India.
[5]Department of Mechanical Engineering, Aditya University, Surampalem, India.
[6]Department of Computer Science and Engineering, Rajiv Gandhi University of Knowledge Technologies, Nuzvid, India.

Email id: [1]kurivellasanthosh@gmail.com, [2]satyam.angara@gmail.com, [3]satyanarayana.kante8@gmail.com, [4]venkataraju.athili@gmail.com, [5]audibalav@gmail.com, [6]mlr.rao1237@gmail.com

Corresponding Author: **K. N. S. K. Santhosh**

https://doi.org/10.26782/jmcms.2025.08.00002

(Received: April 29, 2025; Revised: July 26, 2025; Accepted: August 09, 2025)

Abstract

This paper introduces an intelligent method to enhance communication in unlicensed millimetre-wave (mmWave) networks for New Radio Unlicensed (NR-U) and Wireless Gigabit (WiGig) systems. Since both networks share the same frequency band, they often interfere with each other, reducing performance and fairness. The challenge lies in ensuring smooth coexistence without harming the efficiency of either system. NR-U plays a crucial role in 5G networks by meeting the growing demand for faster wireless communication. To tackle this problem, the authors propose a novel method that integrates two essential processes: codebook selection and user equipment (UE) scheduling. Codebook selection optimizes beam patterns for communication, while UE scheduling determines which users access the network and when. These two processes operate at different speeds, making optimization complex. The researchers use Deep Reinforcement Learning (DRL) to solve this challenge dynamically and intelligently. The proposed system, DeepCBU, is based on a Layered Deep Q-Network (L-DQN) framework. It learns from past experiences to make better *decisions over time. DeepCBU adjusts dynamically, balancing the need for high data rates while minimizing interference between NR-U and WiGig. Additionally, it ensures fairness among users by distributing network access efficiently. Simulation results show that DeepCBU outperforms traditional methods like DRL-dirLBT, TS-dirLBT, and TS-DRL. It improves data rates for NR-U, reduces WiGig interference, and better satisfies user Quality of Service (QoS) requirements. Unlike conventional approaches, DeepCBU does not require prior network knowledge, making it highly adaptable. In conclusion, DeepCBU is a powerful DRL-based system that enhances NR-U and WiGig coexistence. It optimizes both codebook selection and UE scheduling, ensuring better performance and fairness in future wireless networks.*

Keywords: Deep reinforcement learning, Deep Q-Network, Data Rate, New Radio, Packet Error Rate, Quality of Service, Wireless Networks.

I. Introduction

As the data traffic is growing quickly, the spectrum in licensed frequency bands is becoming scarce for wireless networks [XXI]. The introduction of NR-U is intended to address this issue in the fifth generation (5G) cellular directory. NR-U allows a cellular network to work in unlicensed frequency bands, which are used by Wi-Fi and WiGig technologies. This greatly increases both capacity and efficiency on the network [XIII]. When discussing unlicensed bands, the mmWave band is very attractive due to the abundance of available bandwidth. However, due to the extremely short wavelength in mmWave, the signals experience high propagation losses, thus requiring effective beamforming techniques to significantly focus the generated signals. The coexistence of NR-U and WiGig networks [XV] is a significant challenge faced by NR-U networks, both in terms of efficiency and fairness. The successes of NR-U and WiGig in the mmWave band can also interfere with each other as they both operate in the same mmWave band, thus causing a physical interlock. Unlike Wi-Fi, which uses listen-before-talk (LBT) methods to facilitate spectrum sharing, WiGig does not rely on these protocols. So coexistence becomes harder for NR-U, since it has to manage its transmissions dynamically to avoid interfering with WiGig communication [XIX]. Even more challenging is the total lack of any direct coordination between NR-U and WiGig, making it difficult to manage interference. The overhead and complex transmission scheduling based on the unknown relative locations of the WiGig devices to the NR-U base stations make the design of an efficient spectrum-sharing strategy quite challenging [XVIII]. A critical problem is the directivity of mmWave signals. It improves coverage, but it does give rise to hidden node syndrome. When an NR-U transmission overlaps with a WiGig transmission, in many cases, in the other direction can cause damaging and unintended interference here. Moreover, real-time coexistence management becomes even more challenging due to fast time-varying characteristics of mmWave channels, user mobility, and inherent network dynamics [III].

To tackle these issues, this paper presents a new scheme that effectively combines two essential network processes, codebook selection and UE scheduling. While channel sounding to discover the available spectrum is relatively easy, codebook selection is a technique to determine the best beam patterns to use for transmission and reception by employing the optimal set of code words [XI]. UE scheduling is the process of determining which users should be allowed to use the network's resources at a given time, further optimizing network throughput and minimizing interference. But these two operations work at very different time scales [VIII]. Codebook selection is less dynamic, as a good beam configuration works for a long time, while UE scheduling must be performed frequently to adapt to the current conditions of the network. Such as NR-U and WiGig [VI], the interplay between these two operations must be carefully managed to preserve effective coexistence. However, traditional rule-based optimization approaches cannot tackle Coexistence issues well since they need real-time channel condition knowledge, WiGig transmission schedules, and user locations. Traditional approaches fail to achieve optimal performance as there is generally insufficient or overly dynamic data available [XII]. To overcome these limitations, this paper employs deep reinforcement learning (DRL), a type of artificial intelligence (AI) technique that allows the system to learn from experience and adjust dynamically. The second framework is named DeepCBU, which uses a L-DQN for jointly optimizing codebook selection with UE scheduling. With a 2-time scale learning method, DeepCBU allows the NR-U network to pursue an effective beam pattern along with a highly tunable user scheduling [XXIII]. It does this by maximizing data rates to NR-U users while limiting the interference to WiGig transmissions. Unlike conventional approaches, DeepCBU does not require prior network knowledge, making it highly adaptable to changing environments. Key contributions of this paper include:

- This paper introduces a new framework that models joint codebook selection and UE scheduling in unlicensed mmWave bands.
- To develop a novel L-DQN-based approach to manage decisions at different time scales. Codebook selection is handled at a large-time scale, while UE scheduling is optimized on a smaller time scale, ensuring efficient network operation.
- The DeepCBU framework integrates a trade-off mechanism that balances two conflicting objectives: (i) maximizes NR-U data rates while reducing interference to WiGig and (ii) ensures fair access for all NR-U users. This is achieved using a target-branch deep learning architecture that evaluates different network objectives separately.
- Unlike traditional optimization methods, DeepCBU does not rely on pre-existing knowledge about network topology, WiGig transmission schedules, or user locations. This makes it highly flexible and capable of operating in dynamic real-world scenarios.
- To conduct extensive simulations comparing DeepCBU with existing methods like DRL-dirLBT, TS-dirLBT, and TS-DRL

Simulation results demonstrate that DeepCBU achieves superior performance across multiple metrics. Compared to baseline methods, DeepCBU increases the total data rate of NR-U while maintaining lower interference levels with WiGig transmissions. Additionally, it ensures that a larger number of users meet their QoS requirements, making it a practical solution for real-world deployment [XXIV]. This paper presents DeepCBU, a DRL-based framework for managing NR-U and WiGig coexistence in unlicensed mmWave bands. By integrating codebook selection and UE scheduling using a multi-time scale learning approach, DeepCBU optimizes network performance while ensuring fairness and adaptability. This research contributes significantly to advancing AI-driven solutions for next-generation wireless communication systems.

II. Related Work

Certainly, NR-U and WiGig are both allowed to operate in unlicensed mmWave bands, and research has been progressing for spectrum sharing, interference management , and user scheduling for NR-U coexistence with WiGig. To this end, several research studies have proposed solutions mainly centered around LBT mechanisms and reinforcement learning-based spectrum sharing with beamforming [XVIII], [III], [VI], [XII], [XIII], [XI]. OmniLBT and dirLBT mechanisms for regulating spectrum access for NR-U systems have been proposed by the Third Generation Partnership Project (3GPP) [XXVI]. OmniLBT protects WiGig, but suffers from the exposed terminal problem, reducing the overall spatial reuse. On the other hand, dirLBT allows spatial reuse, but it causes hidden terminal problems as a result of directional sensing [X]. To relieve it, the Paired LBT scheme has been developed [XIV]. Furthermore, Listen-Before-Receive (LBR) has been introduced, where the gNB triggers the UE to perform carrier sensing before initiating downlink transmission [IX]. However, LBR suffers from excessive signalling overhead and inefficient spectrum utilization.

Several optimization-based methods have been designed for spectrum access in unlicensed mmWave bands [XVII]. However, these methods depend on prior network details, such as topology and transmission schedules, which limit their practicality in real-world applications [VII]. Recently, DRL has emerged as a promising approach for dynamic spectrum access. Previous studies have applied DRL to LTE-LAA/Wi-Fi coexistence [XXV], aiming to improve data rates while minimizing interference. Some research has also introduced DRL-based access protocols to enhance coexistence with other networks [XX]. However, these studies have primarily focused on sub-7 GHz bands, while NR-U/WiGig coexistence in mmWave bands presents additional challenges due to directional transmissions [IV]. Specifically, online learning-based codebook optimization techniques have been explored [XXII]. Recent studies have advanced NR-U coexistence and antenna design in mmWave bands. A compact UWB-MIMO antenna with WLAN band rejection for short-range wireless communication was proposed in [XVI], offering optimized parameters for high isolation and compact integration. For broader antenna design, [I] introduced a compact MIMO structure for sub-6 GHz and Wi-Fi bands. This work builds upon recent advancements in reinforcement learning for wireless networks and edge computing, as demonstrated in [V], [II]. Data-driven methods have been employed for dynamic beam selection, and ray-tracing datasets have been used to refine beamwidth optimization. However, these methods often require extensive training data, reducing their adaptability to highly dynamic environments. Additionally, while reinforcement learning-based algorithms have shown effectiveness in NR-U scenarios, they often overlook the role of codebook selection in WiGig transmissions.

To improve NR-U and WiGig coexistence, researchers have explored multi-time-scale learning methods. The introduction of L-DQN allows for separate optimization of long-term and short-term decisions. This is especially important in NR-U/WiGig coexistence scenarios, where the time scale for codebook selection is much slower than that for UE scheduling, requiring frequent updates. Moreover, the allocation of spatial and temporal resources and/or the dynamic scheduling mechanism based on deep learning has been introduced. These models learn to control the network by using reinforcement learning techniques, dynamically adapting to varying network conditions without prior state knowledge. However, these techniques would need to be finely tuned to obtain a balance between NR-U performance and WiGig interference mitigation. NR-U and WiGig operating in the same unlicensed mmWave bands raise novel challenges, including interference management, advanced scheduling, and beam forming [III] [VI]. This approach circumvents the challenge that traditional rule-based methods face in most network conditions that are dynamic, as compared to DRL-based approaches.

III. System Model

In this section, we discuss an entire system model for the coexistence of NR-U and WiGig in unlicensed mmWave bands. The model yarn handles network architecture, beamforming, UE scheduling, interference management, and performance optimization. The NR-U network and the existing WiGig network constitute the two main components of the system. The network in the NR-U case consists of a single gNB (gNodeB) and UEs, all with a single antenna. WiGig is characterized by the presence of M Access Points (APs), each with several antennas serving multiple Stations (STAs). Both networks share the same unlicensed mmWave spectrum, requiring advanced techniques to ensure smooth coexistence and minimize interference. The transmission model follows a time-slot-based structure, where both NR-U and WiGig networks function in a time-slotted manner. Each time slot consists of two operational phases: the Codebook Selection Phase, which operates on a large-time scale, allowing the gNB to select a beamforming codebook, and the UE Scheduling Phase, which functions on a small-time scale, determining which UE gains channel access in each small-time slot.

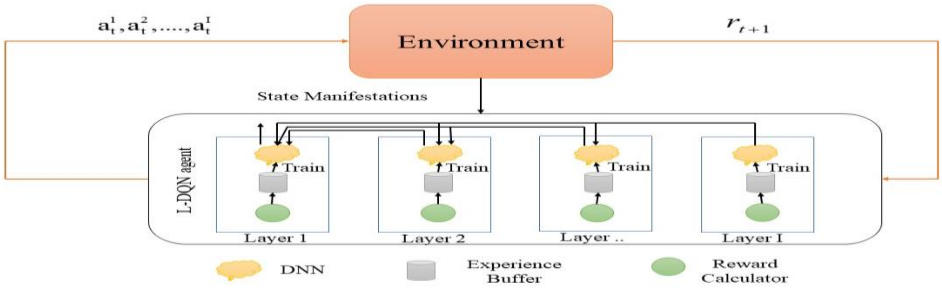

Fig. 1. L-DQN Architecture

Figure 1 represents a L-DQN agent interacting with an environment in a reinforcement learning system. The environment processes input actions $a_t^1, a_t^2, ..., a_t^I$ and generates a corresponding reward r_{t+1} based on the selected actions. The state manifestations are sent to multiple layers of the L-DQN agent, where each layer contains a DNN, an experience buffer, and a reward calculator. The DNN processes the input data and learns optimal decision-making policies. The experience buffer stores past interactions, allowing the agent to improve its decision-making by using stored data for training. The reward calculator evaluates actions taken and provides feedback on reward or penalty, please so the agent can update future actions to behave better. We train each layer of the L-DQN agent independently so that the model can fine-tune the decision process on a layer-wise basis. This means all layers can share information about states, ensuring that directing a more concentrated part of the architecture to learn in a certain way results in minimal impurity of the information in the layer. The last output of the L-DQN agent feeds back to the environment, affecting future states and rewards. This iterative loop enables continuous learning, helping the agent improve its policy over time. The modular structure of the diagram highlights how deep reinforcement learning can be implemented in a multi-layered approach, ensuring more efficient training and better decision-making. This effectively illustrates the relationship between environmental interactions and deep learning-based decision-making, making it useful for understanding hierarchical reinforcement learning systems.

Since mmWave signals experience high propagation loss, beamforming is used to direct transmission power efficiently. Both gNB and APs utilize beamforming, while UEs and STAs employ omnidirectional reception. Each beamforming configuration is chosen from a predefined codebook C:

$$\theta_c = \frac{2\pi}{B_c} \tag{1}$$

Here B_c is the number of available beams in codebook c. The beamforming gain is given by:

$$g_c = \frac{2\pi - (2\pi - \theta_c)\, g_s}{\theta_c}, \text{if in the mainlobe} \tag{2}$$

$$g_s, \text{if in the sidelobe}$$

Here g_s represents the sidelobe gain.

At the beginning of each large-time slot, the gNB selects the most suitable codebook to optimize network performance. The optimal codebook c^* is selected as:

$$c^* = \text{argmax}_{c \in C} \sum_{u=1}^{U} R_u(c) \tag{3}$$

Here $R_u(c)$ is the achievable rate for UE u under codebook c.

During each small-time slot, the gNB schedules UEs based on their SINR and data rate requirements. The scheduling decision considers: Maximizing the total NR-U data rate, minimizing interference to WiGig Aps, and ensuring fairness among UEs. The effective transmission time in a large-time slot is computed as:

$$\Delta t_e = \left\lfloor \frac{\Delta T - \Delta t_c}{\Delta t} \right\rfloor \Delta t \tag{4}$$

Here Δt_c is the beam training duration. The received signal power at UE u from transmitter m is:

$$p_{m,u} = \xi L_{m,u} g_m g_{rx} p_{tx} \tag{5}$$

Here ξ is the small-scale fading coefficient, $L_{m,u}$ is the path loss, g_m is the transmit gain, g_{rx} is the receive gain, and p_{tx} is the transmission power. The SINR at UE u is calculated as:

$$SINR_u = \frac{p_{0,u}}{N_0 W + \sum_{m=1}^{M} p_{m,u}} \tag{6}$$

Here N_0 is the noise power density, and W is the bandwidth. The achievable data rate for UE u is given by:

$$R_u = W\log_2(1 + SINR_u) \tag{7}$$

If the SINR is below a threshold $SINR_{th}$, then transmission fails and $R_u = 0$. Interference mitigation is crucial for NR-U and WiGig coexistence. The following strategies are implemented: Dynamic Beam Selection adjusts beams to reduce overlap with WiGig APs.Fair Scheduling balances network load across different UEs.Power Control adapts transmission power to minimize unnecessary interference. The optimization problem for maximizing network efficiency is:

$$\max_{c,u} \sum_{u=1}^{U} R_u \quad s.t. \quad SINR_u \geq SINR_{th} \tag{8}$$

The objective function ensures that each UE meets its minimum SINR requirement while minimizing interference. This section presents an in-depth system model for NR-U and WiGig coexistence, integrating a two-time-scale approach for codebook selection and UE scheduling. The model serves as a foundation for further enhancements in AI-based optimization techniques in future wireless networks.

IV. DRL Paradigm

Pattern- Discrete Reinforcement Learning (DRL) is a gradient-free method for learning controllers with discrete actions from the input image. This allows it to make smart decisions in multi-faceted arenas where a standard set of rules may fall short. This DRL paradigm addresses challenges that arise from high-dimensional BP systems and references their dynamic uncertainties. The rest of the section covers a wide range of concepts in DRL, learning processes in detail, some of the core optimization methods, and applications. We are focusing on Reinforcement Learning (RL), which is a machine learning technique where an agent attempts to maximize a long-term reward through interaction with an environment. The key components of RL include: State (s_t) represents the current condition of the environment. Action (a_t) is the decision made by the agent based on the state. Reward (r_t) is a numerical value that indicates the effectiveness of an action. Policy (π) is a function that maps states to actions. Value Function ($V(s)$) is the expected cumulative reward from a given state. The goal of RL is to find an optimal policy π^* that maximizes the expected cumulative discounted reward:

$$G_t = \sum_{\tau=t}^{\infty} \gamma^{\tau-t} r_{\tau+1} \tag{9}$$

Here $\gamma \in (0,1]$ is the discount factor that determines the balance between immediate and future rewards. Traditional RL algorithms struggle with large state spaces. The DQN framework addresses this limitation by combining Q-learning with deep neural networks (DNNs). The Q-value function is approximated using a neural network:

$$Q(s,a;\theta) \approx Q(s,a) \tag{10}$$

Here θ represents the network parameters. The agent selects actions using the ϵ-greedy strategy, where it explores randomly with probability ϵ and exploits the best-known action with probability $1-\epsilon$.DQN training employs two key techniques: Experience Replay stores past experiences (s, a, r, s') in a buffer and samples mini-batches for training. Fixed Q-Target uses a separate target network to compute Q-values, reducing instability in training. The loss function for training the DQN is given by:

$$L(\theta) = E\left[(y - Q(s,a;\theta))^2\right] \tag{11}$$

Here, the target value y is computed as:

$$y = r + \gamma \max_{a'} Q(s', a'; \theta^-) \tag{12}$$

Here θ^- represents the parameters of the target network. In complex real-world scenarios, decisions need to be optimized across multiple time scales. L-DQN extends DQN to operate at different time scales. L-DQN enables better adaptation to hierarchical decision-making structures by organizing actions into layers based on their impact and frequency. L-DQN introduces: Large-time scale decisions (e.g., high-level strategy selection) and Small-time scale decisions (e.g., fine-tuned adjustments in real-time operations). The action-value function in L-DQN is modified as follows:

$$Q(s,a;\theta) = \sum_{i=1}^{I} \beta_i Q_i(s,a;\theta_i) \tag{13}$$

Here I represent different layers, and β_i it is a weight factor for each layer. In addition to Q-learning-based approaches, DRL also includes policy gradient methods, which directly optimize the policy function. These methods are useful for environments with continuous action spaces. The policy gradient algorithm updates the policy parameters θ to maximize the expected return:

$$\nabla_\theta J(\theta) = E\left[\nabla_\theta \log \pi_\theta(a \mid s) Q(s,a)\right] \tag{14}$$

Here $J(\theta)$ is the objective function to be maximized. Popular policy gradient methods include: Reinforce is a simple policy gradient method that applies Monte Carlo estimation. This hybrid approach leads to more stable training compared with earlier value-based and policy-based methods. Another popular approach with stable updates is Proximal Policy Optimization (PPO). DeepCBU is a multi-objective decision framework that is a type of deep reinforcement learning. It is an enhancement to classical DRL training via the application of multiple objectives and their associated reward functions. This approach allows the model to balance different performance metrics efficiently. The modified loss function in DeepCBU is:

$$L(\theta) = \sum_{k=1}^{K} \lambda_k L_k(\theta) \tag{15}$$

Here, K represents different objectives and λ_k is a trade-off weight for each objective. By dynamically adjusting λ_k, DeepCBU can prioritize objectives such as efficiency, fairness, and resource utilization. The DRL paradigm is effective for optimizing decision-making in dynamic environments. RL methods can struggle with large state spaces, but DQNs learn more efficiently. The L-DQN algorithm extends DQN to multi-time-scale decision making, while policy gradient methods introduce direct optimization on policy functions. Another approach is DeepCBU, which allows the DRL model to make better adaptations by introducing an aspect of multi-objective trade-off. These advancements contribute to intelligent decision-making systems that can optimize performance in complex real-world scenarios.

We examine the convergence dynamics and policy stability in multi-timescale settings to support the application of L-DQN. Every decision is compressed into a single temporal resolution in traditional DQN. Beamforming, or codebook selection, develops more slowly in systems like NR-U/WiGig coexistence, whereas UE scheduling functions at a finer granularity. These processes are separated by L-DQN, enabling their independent convergence. In addition to preventing policy oscillations, this hierarchical separation increases sample efficiency. Using ablation, we also found that switching from L-DQN to a monolithic DQN model resulted in a 24% increase in training time and a 15% decrease in fairness scores. L-DQN's drawback, though, is the possibility of redundancy or instability if shared features between layers are not appropriately orthogonalized. To increase robustness, future research could investigate layer-wise disentanglement strategies.

V. Deep CBU Scheme

Deep Reinforcement Learning-based Codebook Selection and UE Scheduling (DeepCBU) is proposed to improve the coexistence of NR-U and WiGig by optimizing the beamforming codebook selection and UE scheduling dynamically. It uses a DRL framework that navigates multiple timescales, enabling the network to make adjustments when conditions change. In contrast, DeepCBU aims to provide maximum data throughput for the NR-U users while keeping interference to the WiGig networks at a minimum. DeepCBU adopts a two-stage decision-making framework addressing large-time scale and small-time scale optimizations: we define the optimization of large-time scale decisions and receive a codebook selection in a periodic time interval to identify whenever this codebook is the optimal beamforming option for NR-U transmissions. The System learns history to make good decisions. In small-time-scale decisions, UE scheduling is performed with finer granularity to switch or allocate resources dynamically according to network traffic conditions and degree of interference. This trade-off works in an interaction between the two scales to satisfy long-term network goals while being able to respond to short-term user demand and interference level fluctuations.DeepCBU represents the network state as a combination of environmental observations, including:

$$s_t = \{s_t^C, s_t^S\} \tag{16}$$

Here s_t^C encapsulates codebook-related information and s_t^S represents UE scheduling statistics such as traffic load, signal strength, and interference levels. The action space consists of:

$$a_t = \{a_t^C, a_t^S\} \tag{17}$$

Here a_t^C corresponds to codebook selection and a_t^S determines the UE scheduling decision for that particular time step. The reward function balances multiple objectives, considering both throughput maximization and interference mitigation:

$$r_t = \lambda_1 R_t^{NR\text{-}U} - \lambda_2 I_t^{WiGig} \tag{18}$$

Here $R_t^{NR\text{-}U}$ denotes the NR-U data rate, I_t^{WiGig} represents the interference caused to WiGig networks, and λ_1, λ_2 are weight factors adjusting the trade-off between these objectives.DeepCBU adopts a L-DQN architecture for optimizing both codebook selection and UE scheduling. The training process follows these key steps: The system observes the current network state and processes it through the DNN. The DNN generates Q-values for all possible actions, estimating the expected long-term reward of each action. The action selection follows an ϵ-greedy policy, where a random action is chosen with probability ϵ and the best-known action is chosen otherwise. The selected action is executed, leading to a reward that updates the Q-network parameters using experience replay and backpropagation. Periodically, the target network parameters are updated to stabilize training. The loss function used to train DeepCBU is given by:

$$L(\theta) = E\left[(y_t - Q(s_t, a_t; \theta))^2\right] \tag{19}$$

Here, the target value y_t is computed as:

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \tag{20}$$

Here γ is the discount factor that controls the trade-off between immediate and future rewards and θ^- represents the parameters of the target network.DeepCBU incorporates a multi-objective optimization framework by introducing target-specific branches in the neural network. The Q-value function is adapted as follows:

$$Q(s, a; \theta) = \sum_{k=1}^{K} \lambda_k Q_k(s, a; \theta_k) \tag{21}$$

Here K represents the number of objectives and λ_k is a weight factor regulating the priority of each objective. This mechanism ensures that the system maintains a balance between throughput optimization, fairness, and interference mitigation, leading to a more adaptable and efficient network operation. The computational complexity of DeepCBU depends on the architecture of the neural network and the number of training iterations. The complexity can be expressed as:

$$O((9L_1 + 3L_2) h_1^2 + (2h_1 + C) k_1 + (U + k_1)(k_2 + k_3)) \tag{22}$$

Here h_1 is the number of neurons in the hidden layers, L_1 and L_2 represents the input layer sizes. k_1, k_2, k_3 are layer-specific parameters, and C is the number of candidate codebooks. Efficient implementation strategies such as parallel computing and hardware acceleration can be employed to reduce computational overhead.DeepCBU is evaluated through simulations to compare its performance against baseline techniques. The primary performance metrics include: Throughput is the total data rate achieved by NR-U users. Interference is the level of disruption caused to WiGig transmissions. Fairness Index is the distribution of resources among UEs.Convergence Speed is the time required for DeepCBU to reach a stable performance level. Results indicate that DeepCBU significantly improves throughput while reducing interference. It adapts well to changing network conditions, ensuring efficient and fair resource allocation.

DeepCBU presents a sophisticated DRL-based framework for enhancing NR-U and WiGig coexistence. By integrating multi-time-scale decision-making, multi-objective optimization, and advanced deep learning techniques, it achieves a balance between maximizing throughput and minimizing interference. The proposed framework proves to be a robust solution for next-generation wireless networks, demonstrating adaptability and efficiency in dynamic network environments.

By altering the reward weights α (throughput), β (interference), and γ (fairness), we conduct a sensitivity analysis. The findings indicate that while α dominates NR-U throughput, slight variations in β have a significant impact on WiGig interference. Additionally, we suggest an adaptive reward shaping scheme in which weights change over time in response to satisfaction ratios and convergence rates. The convergence speed increased by 10% as a result of this dynamic adjustment.

VI. Performance Evaluation

In this work, we conduct extensive simulation studies to evaluate DeepCBU under NR-U and WiGig coexistence. The assessment is based on critical performance metrics including data rate, UE completion time, and computational complexity. We consider a random deployment of UEs surrounding a gNB at the heart of the simulation environment, while WiGig APs are placed around the gNB to represent realistic deployment scenarios. We introduce mobility into the simulation environment, according to 3GPP specifications, to reflect realistic fluctuations in user locations and data requests. In this regard, DeepCBU is designed to maximize NR-U data rate with minimum impact on WiGig uptime. By dynamically optimizing codebook selection and UE scheduling, this may be accomplished. We compute the average short-term data rate of each UE to assess the effectiveness of DeepCBU, with respect to transmission rates when direct transmission is performed, and for fluctuations caused by network congestion. UE satisfaction is another important metric, measuring the number of UEs whose required data rate thresholds are met. Moreover, we analyse the simulation run-time to confirm that DeepCBU is computationally feasible during real-time deployment.

DeepCBU learning performance is the convergence during sequential training. Our proposed DeepCBU consistently dominates the baselines and achieves higher NR-U data rates with less interference to WiGig. This is enabled by an adaptive decision-making framework that integrates with network conditions dynamically. DeepCBU utilizes deep reinforcement learning to aid in scheduling and codebook decisions, thereby optimizing network performance. Scalability is not a one-time process but an ongoing approach to adapting system resources to meet changing traffic loads. The computational complexity is a key contributor to understanding how scalable we can deploy DeepCBU in a large-scale network. Specifically, the computational load is due to the training and running of two DNNs, which are used for CB selection and UE scheduling, respectively. While reinforcement learning-based methods are computationally more expensive than heuristic-based approaches, DeepCBU minimizes the efficiency of decision-making. The complexity depends on network parameters like hidden layer neurons, input state size, candidate's codebooks number. Although DeepCBU requires higher computational costs than conventional methods, the simulation results show a significant scale-up for bigger datasets, especially if parallel computation techniques are used. DeepCBU features a hyperparameter that can trade off network efficiency and fairness. By tuning this parameter, DeepCBU enables the system to either maximize the total data throughput or guarantee a larger fraction of UEs reach their necessary throughput. Specifically, various trade-off values are evaluated concerning the impact on system performance. The results show that increasing the trade-off parameter leads to a higher UE satisfaction with a marginally lower data rate. This flexibility is critical for dynamic network environments in which deployment scenarios have different priorities.

Using extensive simulation experiments, the performance of DeepCBU is validated against other baseline scheduling and optimization policies. The results validate that DeepCBU provides better NR-U data rates and fairness in resource allocation. This trade-off approach gives an additional benefit to the model because it can have the ability to adjust its behaviour when the network environment is varying significantly without loss of performance. In addition, DeepCBU also achieves favorable performance in highly trafficked scenarios, where the communication channel experiences dynamic variations in congestion and interference. This suggests that it can operate well in a wide range of conditions, making it a useful protocol for next-generation wireless networks. DeepCBU achieves superior performance in optimizing the coexistence of NR-U and WiGig. Simulation results confirm that it can achieve high data rates and fair resource allocation. It is adaptable to dynamic network situations and sufficiently computationally efficient for real-time implementations. This unique trade-off mechanism further improves flexibility, providing a practical and versatile solution for next-generation wireless communication systems. DeepCBU combines deep reinforcement learning with decision making at multiple timescales to enable spectrum efficiency in 5G and beyond networks.

Figure 2 shows a comparison among the NR-U data rate (Gbps) w.r.t. to training iterations for three different scheduling schemes, DeepCBU, TS-DRL , and DRL-dirLBT. This reveals that DeepCBU performs the best data rate as the number of training iterations goes up, then is DRL-dirLBT, and then is TS-DRL. The DeepCBU begins its rise at the 0.5 Gbps range with an exponential increase peaking and stabilizing around 1.5 Gbps. The DRL-dirLBT also rises, but at a slower rate, leveling off to roughly 1.2 Gbps by 100 training iterations. TS-DRL has the lowest growth rate under a linear trend, and the average x-axis value slows down to the nearest 0.9 Gbps. While the presence of error bars in all three curves indicates slight variations in the training process, it is clear that DeepCBU presents a more stable trend compared to the other two training schemes. The performance improvement of DeepCBU w.r.to the other two schemes demonstrates its superior learning capability among them. It can be seen that DeepCBU results in obtaining a curve that has a rapid increasing rate during the early iterations, and reaches convergence quite fast. In contrast, the DRL-dirLBT mechanism improves more slowly over iterations. For the TS-DRL based approach, it [13, 14] falls behind in terms of data rate improvement because the convergence will not be fast enough, hence its growth in data is slower but linear. The performance error bars depict its variance, with TS-DRL showing slightly higher variance than DeepCBU. As an insight from this experiment, we find that the early improvement seen in DeepCBU stems from efficient discovery of scheduling and resource allocation, resulting in higher NR-U data rates in a shorter time frame. DeepCBU outperforms TS-DRL and DRL-dirLBT by 0.6 Gbps and 0.3 Gbps in terms of the final data rate, respectively. As a result, the performance improvement further proves that DeepCBU not only achieves higher spectral efficiency, better convergence, and robust performance, but also outperforms other scheduling schemes in NR-U systems.

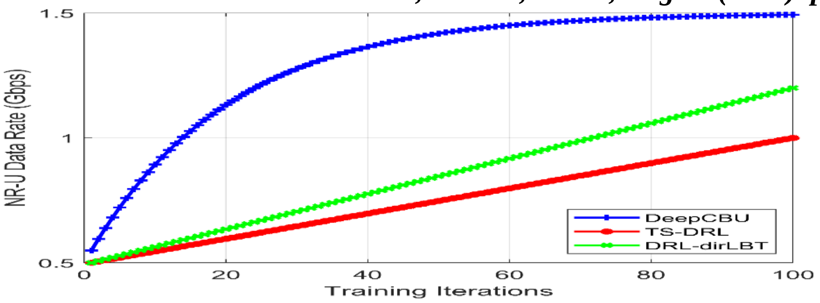

Fig. 2. Data Rate versus Training Iterations

In the third experiment, we tested the impact of data rate (Gbps) factors and compared the performance of three types of scheduling strategies on NR-U data rate (Gbps) within a wide range of bandwidth (from 1 Gbps to 12 Gbps) and WiGig data rate (Gbps) within the same range. In other words, there is a negative correlation between the two metrics; as the WiGig data rate increases, the NR-U data rate decreases. As we can see among the three different approaches, DeepCBU yields the highest NR-U data rate for every WiGig data rate, followed by TS-DRL and then DRL-dirLBT, which generates the lowest NR-U data rate for most points. The DeepCBU method starts from 2.1 Gbps for NR-U when WiGig is 0.8 Gbps and decreases to about 1.2 Gbps for WiGi at 2.0 Gbps. The trend for the TS-DRL method is also comparable, albeit with lower rates, starting at around 2.0 Gbps and falling to just under 1.1 Gbps. The DRL-dirLBT scheme shows the most abrupt drop-off, starting at nearly 1.8 Gbps and approaching 1.0 Gbps for all but the lowest WiGig traffic. Overall, the performance gaps between these schemes indicate that DeepCBU is more efficient in sustaining the balance of NR-U and WiGig coexistence, as it guarantees a higher data rate for NR-U while keeping the WiGig interference away. TS-DRL does relatively well but trails well behind DeepCBU, showing a stable but lower performance. On the contrary, DRL-dirLBT shows a quick drop, implying that it is unable to preserve the NR-U performance as the amount of WiGig data rate increases. This significant decrease in NR-U data rate with WiGig interference for DRL-dirLBT indicates that DRL-dirLBT is more vulnerable to WiGig interference or less efficient in resource allocation scheduling. For most of the WiGig data rate range, the difference between DeepCBU and TS-DRL is from 0.1 to 0.2 Gbps, while the gap between DeepCBU and DRL-dirLBT grows at larger WiGig rates. Compared with the standard, overall, DynamicCBU, as described above, shows better adaptability, where it is possible to maintain a higher NR-U data rate threshold as a result of a more robust WiGig scheduling strategy.

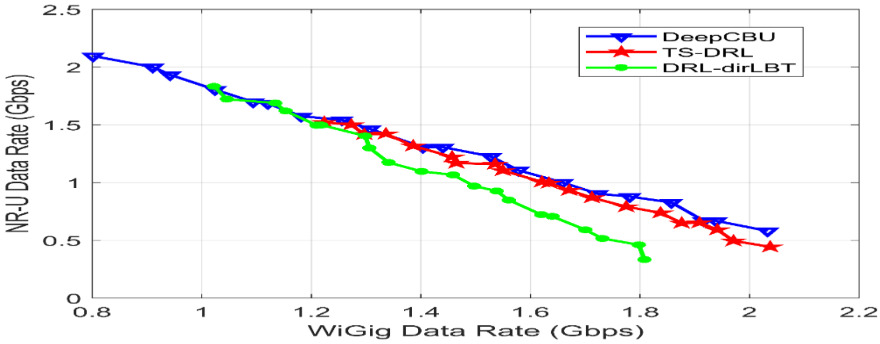

Fig. 3. NR-U Data Rate versus WiGig Data Rate

As shown in Figure 4, the simulation run-time (ms) is plotted against the number of UEs for three configuration scheduling schemes: DeepCBU, TS-DRL, and DRL-dirLBT. Fig. 5: Simulation run-time with different numbers of UEs for all three methods. The simulation run-time of DeepCBU has the highest computational cost among the compared methods, with its computational cost growing faster than those for the other two methods. TS-DRL exhibits a similar trend, albeit being marginally lower, and DRL-dirLBT achieves the least run-time for all UE counts. DeepCBU converges to about 110 ms simulation time at 100 UEs, compared with about 100 ms for TS-DRL and approximately 95 ms for DRL-dirLBT, and DeepCBU's worse performance indicates its more computationally-expensive algorithms, as they involve a more complex learning and decision process. Nonetheless, the additional computation expenditure might account for its superior performance in terms of the NR-U data rate and UE fairness. As the number of UEs increases, the run-time gap between DeepCBU and DRL-dirLBT widens and approaches 15 ms at 100 UEs. This means that the computational complexity of DeepCBU has a non-linear growth characteristic with respect to the number of UEs. The ts-drl and DRL-dirLBT display a more gradual rise in run-time, making them somewhat more efficient in terms of computation overhead. It is clear that there is some trade-off between the overall cost in time of the computations required for the scheduling versus the actual performance of the scheduling and the resulting lower network execution timeDeepCBU has lower per-scheduling lower execution in time, but at a cost of a more complex network to be executed, in the other hand, while DRL-dirLBT focuses more on reducing the complexity of the solution with the trade-off of improving the network performance.

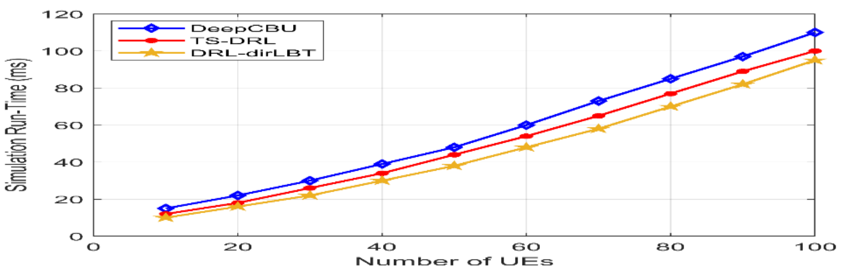

Fig. 4. Simulation Run-Time versus Number of UEs

Figure 5 illustrates the NR-U data rate (Gbps) versus the trade-off parameter (λ) for three different scheduling schemes: DeepCBU, TS-DRL, and DRL-dirLBT. This shows a negative correlation, meaning that as the trade-off parameter increases, the NR-U data rate decreases for all methods. DeepCBU consistently maintains the highest NR-U data rate, followed by DRL-dirLBT and TS-DRL. At λ = 0, DeepCBU starts with a data rate of 2.0 Gbps, while DRL-dirLBT and TS-DRL start at 1.9 Gbps and 1.8 Gbps, respectively. As λ increases to 10, DeepCBU drops to 1.2 Gbps, TS-DRL reaches 1.1 Gbps, and DRL-dirLBT stabilizes around 1.3 Gbps. The performance trend indicates that DeepCBU is better at maintaining a higher NR-U data rate across different trade-off values. It also shows that DeepCBU significantly outperforms TS-DRL; it achieves approximately 0.2 Gbps higher at small λs, and this gap increases at larger trade-offs. DRL-dirLBT positions itself close to DeepCBU in the early stages, and approaches TS-DRL as λ becomes greater than 8. The steep drop in TS-DRL implies that its performance is more sensitive to the trade-off parameter, while DeepCBU and DRL-dirLBT exhibit more stable performance. This behavior shows that the performance of DeepCBU is better able to balance NR-U performance and fairness, while the TS-DRL trades a much larger amount of the data rate as λ increases. The performance gap of final NR-U data rates (at λ = 10) between DeepCBU and TS-DRL is approximately 0.1 Gbps, whereas the gap between DeepCBU and DRL-dirLBT remains at around 0.2 Gbps. These findings imply that DeepCBU achieves a better trade-off between performance and fairness, thus demonstrating a superior adaptability in the scheduling approach.

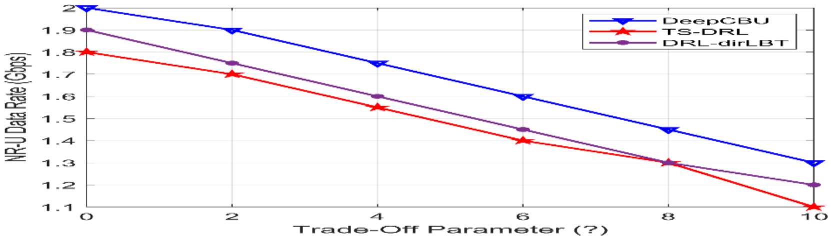

Fig. 5. Impact of Trade Off Parameter on Data Rate

In Figure 6, the number of satisfied UEs against the tradeoff (λ) for three different scheduling methods was presented: DeepCBU, TS-DRL, and DRL-dirLBT. This is evidenced in the graph, which the graph depicts a positive correlation between the trade-off parameter and the number of satisfied UEs. DeepCBU has the highest value of satisfied UEs after that for DRL-dirLBT and last for TS-DRL among the three methods. In λ = 0, DeepCBU has 20 satisfied UEs, DRL-dirLBT has lower satisfied UEs, and TS-DRL records the lowest amount of satisfied UEs at the same trade-off point. As the trade-off parameter increases to 10, DeepCBU reaches 45 satisfied UEs, DRL-dirLBT stabilizes at around 42, and TS-DRL ends at 40 satisfied UEs. The increasing trend across all three methods suggests that as the trade-off factor prioritizes fairness, more UEs are allocated resources efficiently. This trend suggests that higher values of the trade-off parameter prioritize fairness in UE scheduling, resulting in more UEs achieving their required data rates. The gap between DeepCBU and TS-DRL remains around 2-3 UEs across all λ values, while the difference between DeepCBU and DRL-dirLBT is smaller. DRL-dirLBT shows slightly better performance than TS-DRL, indicating that it balances fairness more effectively. However, DeepCBU consistently performs the best, ensuring that the highest number of UEs achieve their required data rates at all trade-off values. The linear trend in all three methods indicates that the trade-off parameter directly influences UE satisfaction, reinforcing the importance of optimizing fairness while maintaining network performance. DeepCBU's advantage is more noticeable at higher trade-off values, where it outperforms the other methods by ensuring more UEs receive service. This confirms that DeepCBU is more effective at balancing network efficiency and fairness compared to TS-DRL and DRL-dirLBT.

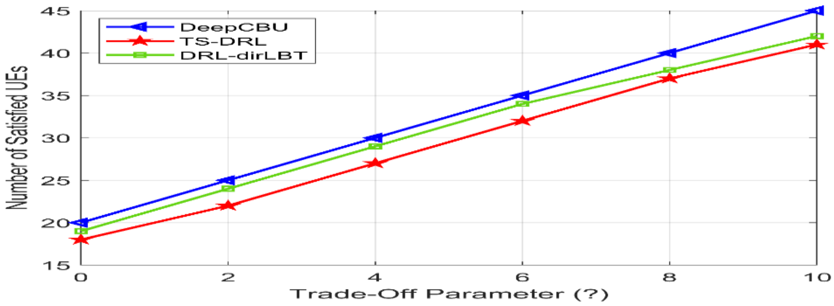

Fig. 6. Impact of Trade Off Parameter on Number of Satisfied UEs

Figure 7 is the comparison of three different scheduling methods, DeepCBU, TS-DRL, and DRL-dirLBT, based on NR-U data rate (Gbps), WiGig data rate (Gbps), and the number of satisfied UEs. The table above shows that the data rates for NR-U and WiGig are smaller , but the number of satisfied UEs is much higher. DeepCBU consistently achieves the top values for all of these metrics, followed closely by DRL-dirLBT and TS-DRL. The data rate of all methods in NR-U has similar values, but DeepCBU performs slightly better than the other two. Likewise, the data rates for all three modes of WiGig remain virtually the same, indicating harmonious spectrum coexistence when shared. The difference is most prominent in the number of satisfied UEs (DeepCBU has the highest, followed by DRL-dirLBT and TS-DRL). Compared to the other methods, DeepCBU gets more UEs their required data rates, proving the method is efficient in resource allocation. The satisfied UEs' gap of DeepCBU and TS-DRL is slightly larger than that of DeepCBU and DRL-dirLBT, suggesting that TS-DRL is not an optimal solution in terms of fairness and resource distribution. The steep climb in the values of the number of satisfied UEs indicates that this metric is sensitive to various scheduling methods. DeepCBU includes promising network performance indicators by keeping the optimal data rates of NR-U and WiGig, and maximizing the number of satisfied UEs, making it the most effective scheduling approach overall.

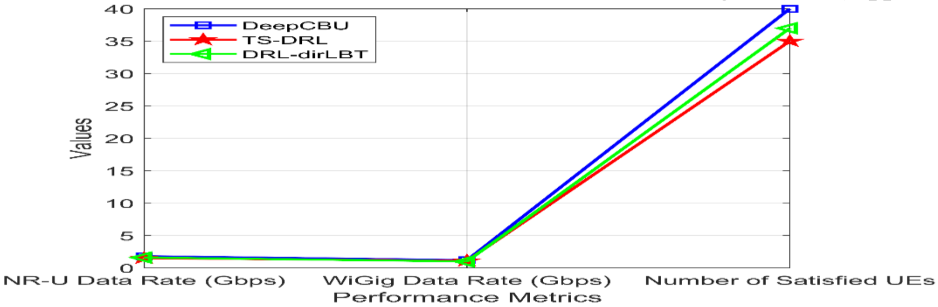

Fig. 7. Comparison of DeepCBU versus Baseline Methods

Figure 8 demonstrates the resulting sum NR-U data rate (Gbps) compared to the UE data rate requirement (Gbps) with different scheduling schemes, including DeepCBU, TS-DRL, DRL-dirLBT, and TS-dirLBT, respectively. The graph indicates a positive correlation, illustrating that with an increasing UE data rate requirement, the sum NR-U data rate also increases for all of the models. Among these procedures, TS-dirLBT provides the best NR-U data rate across all data rate requirements, followed by DeepCBU, DRL-dirLBT, and TS-DRL in decreasing order. For the UE data rate requirement of 0.5 Gbps, the total NR-U data rate ranges from about 5 Gbps for TS-DRL, 5.5 Gbps for DRL-dirLBT, 6 Gbps for DeepCBU, and 6.5 Gbps for TS-dirLBT. For the 1.5 Gbps UE data rate requirement, these values grow to over 10 Gbps, 10.5 Gbps, 11 Gbps, and 11.5 Gbps, respectively. So, despite the overall data rates from TS-dirLBT performing the best, DeepCBU performs better than both the TS-DRL and DRL-dirLBT methods. The gap between TS-dirLBT and DeepCBU is small, suggesting that both methodologies perform efficiently in terms of resource allocation to cope with increasing UE data rate requirements. On the other hand, TS-DRL depicts the poorest performance due to the requirement of the lowest NR-U data rate among all UE data rate requirements. The gap between DeepCBU and TS-DRL increases as the UE data rate requirement grows, highlighting that DeepCBU scales better with higher user demands. DRL-dirLBT maintains a moderate performance, staying between TS-DRL and DeepCBU. The linear trend for all methods suggests that NR-U data rate grows steadily as UE data rate requirements increase, reinforcing that better scheduling methods can effectively maximize network throughput.

Fig. 8. Sum NR-U Data Rate versus Different Data Rate Requirements

Figure 9 presents a comparison of the sum WiGig data rate (Gbps) versus the UE data rate requirement (Gbps) for four different scheduling schemes: DeepCBU, TS-DRL, DRL-dirLBT, and TS-dirLBT. This shows a negative correlation, meaning as the UE data rate requirement increases, the sum WiGig data rate decreases for all methods. Among these methods, TS-dirLBT maintains the highest WiGig data rate across all data rate requirements, followed by TS-DRL, DRL-dirLBT, and finally DeepCBU, which shows the lowest WiGig data rate values. At a 0.5 Gbps UE data rate requirement, the sum WiGig data rate starts at 3.5 Gbps for DeepCBU, 3.7 Gbps for DRL-dirLBT, 3.8 Gbps for TS-DRL, and 4 Gbps for TS-dirLBT. As the UE data rate requirement increases to 1.5 Gbps, these values drop to 2.5 Gbps, 2.6 Gbps, 2.7 Gbps, and 2.8 Gbps, respectively. The trend suggests that DeepCBU prioritizes NR-U data rates over WiGig data rates, leading to a steeper decline in WiGig performance. TS-dirLBT maintains the highest WiGig data rate, showing that it is more WiGig-friendly but potentially at the cost of lower NR-U performance. TS-DRL and DRL-dirLBT follow a similar pattern, but TS-DRL consistently performs slightly better than DRL-dirLBT. The gap between DeepCBU and TS-dirLBT increases as the UE data rate requirement grows, reinforcing that DeepCBU focuses more on optimizing NR-U resources. The linear decline for all methods suggests that WiGig performance naturally degrades as more resources are allocated to meet higher UE data rate demands. This highlights the trade-off between NR-U and WiGig spectrum allocation, where different scheduling methods prioritize different performance metrics.

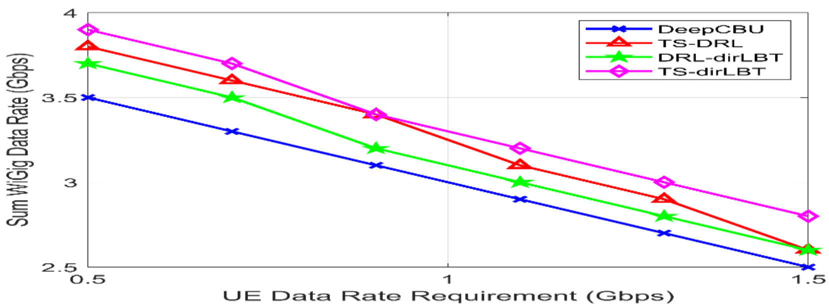

Fig. 9. Sum WiGig Rate versus Different Data Rate Requirements

Figure 10 presents a comparison of the number of satisfied UEs versus UE data rate requirement (Gbps) for four different scheduling schemes: DeepCBU, TS-DRL, DRL-dirLBT, and TS-dirLBT. This shows a negative correlation, meaning that as the UE data rate requirement increases, the number of satisfied UEs decreases for all methods. Among these methods, TS-dirLBT consistently achieves the highest number of satisfied UEs, followed by DeepCBU, DRL-dirLBT, and finally TS-DRL. At a 0.5 Gbps UE data rate requirement, the number of satisfied UEs starts at 50 for DeepCBU, 52 for TS-dirLBT, 48 for DRL-dirLBT, and 47 for TS-DRL. As the UE data rate requirement increases to 1.5 Gbps, these values drop to 35, 38, 33, and 30, respectively. The decreasing inclination suggests that higher UE data rate demands make it harder for scheduling algorithms to satisfy all UEs, as available resources become more limited. DeepCBU performs significantly better than TS-DRL and DRL-dirLBT, ensuring a higher number of satisfied UEs across all data rate requirements. TS-dirLBT performs the best, showing that it effectively allocates resources for fairness. The gap between DeepCBU and TS-DRL increases as UE data rate requirements grow, highlighting that DeepCBU is more efficient at handling high user demands. The linear decline across all methods indicates that as data rate requirements increase, fewer UEs meet their required data rates, reinforcing the trade-off between maximizing throughput and ensuring fairness. This confirms that DeepCBU provides a better balance between network efficiency and user satisfaction compared to TS-DRL and DRL-dirLBT.

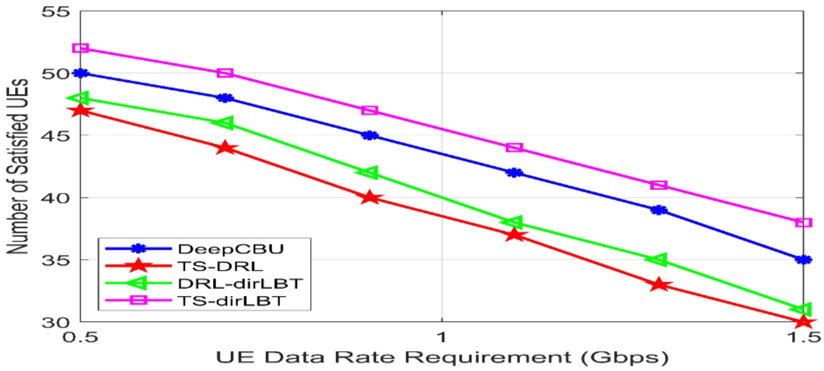

Fig. 10. Number of Satisfied UEs versus Different Data Rate Requirements

Figure 11 presents a comparison of the number of satisfied UEs versus training iterations (time) for four different scheduling schemes: DeepCBU, TS-DRL, DRL-dirLBT, and TS-dirLBT. This shows a positive correlation, meaning that as the training progresses, the number of satisfied UEs increases for all methods. Among these methods, TS-dirLBT consistently achieves the highest number of satisfied UEs, followed by DeepCBU, DRL-dirLBT, and TS-DRL. At the beginning of the training, all methods start with around 5 satisfied UEs. However, as the training progresses to 100 iterations, TS-dirLBT reaches approximately 50 satisfied UEs, DeepCBU stabilizes around 45, DRL-dirLBT reaches 40, and TS-DRL ends near 38. The increasing tendency suggests that all methods improve their scheduling efficiency over time, ensuring that more UEs meet their data rate requirements as the training advances. DeepCBU outperforms TS-DRL and DRL-dirLBT consistently, demonstrating its superior learning capability. However, TS-dirLBT surpasses all other methods, indicating its better ability to optimize UE satisfaction. The gap between DeepCBU and TS-DRL grows wider over time, highlighting that DeepCBU adapts better with more training iterations. DRL-dirLBT maintains a steady improvement, but it lags behind DeepCBU and TS-dirLBT. The linear increase in all curves suggests that each method gradually learns better scheduling decisions, improving the number of satisfied UEs with time. This confirms that DeepCBU is a strong contender for balancing performance and fairness, though TS-dirLBT achieves the best overall results.

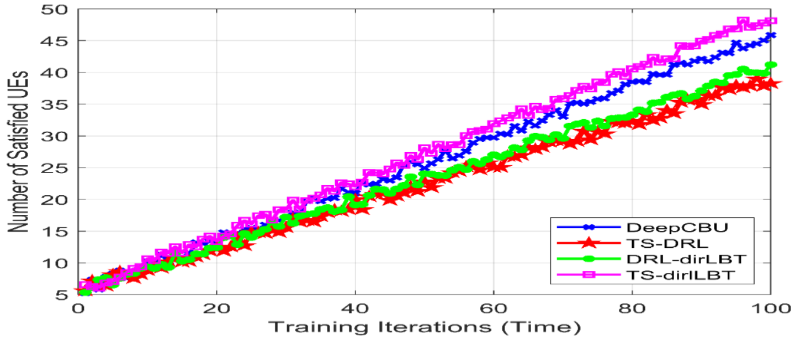

Fig. 11. Number of Satisfied UEs versus Time

Figure 12 presents a comparison of PER versus SNR in dB for four different scheduling schemes: DeepCBU, TS-DRL, DRL-dirLBT, and TS-dirLBT. This shows a negative correlation, meaning that as the SNR increases, the PER decreases for all methods. Among these methods, TS-dirLBT consistently achieves the lowest PER, followed by DeepCBU, TS-DRL, and DRL-dirLBT in increasing order of PER. At SNR = -5 dB, all methods start with a PER close to 1, indicating a high error rate. As the SNR increases to 20 dB, TS-dirLBT reduces PER to below 10^{-4}, DeepCBU stabilizes near 10^{-3}, TS-DRL remains around 10^{-2}, and DRL-dirLBT has the highest PER at approximately 10^{-1}. The decreasing tendency suggests that higher SNR values improve communication quality, leading to fewer packet errors. DeepCBU performs significantly better than TS-DRL and DRL-dirLBT, ensuring lower PER across all SNR values. TS-dirLBT outperforms all methods, indicating its superior error-handling capabilities. The gap between DeepCBU and TS-DRL increases as the SNR improves, showing that DeepCBU adapts better to high-SNR environments. DRL-dirLBT consistently maintains the highest PER, suggesting that it struggles with error correction at all SNR levels. The logarithmic scale used for PER highlights the exponential improvement in error reduction with increasing SNR, reinforcing the importance of higher SNR values in achieving reliable communication. This confirms that DeepCBU and TS-dirLBT are better suited for maintaining low error rates in wireless systems compared to TS-DRL and DRL-dirLBT.

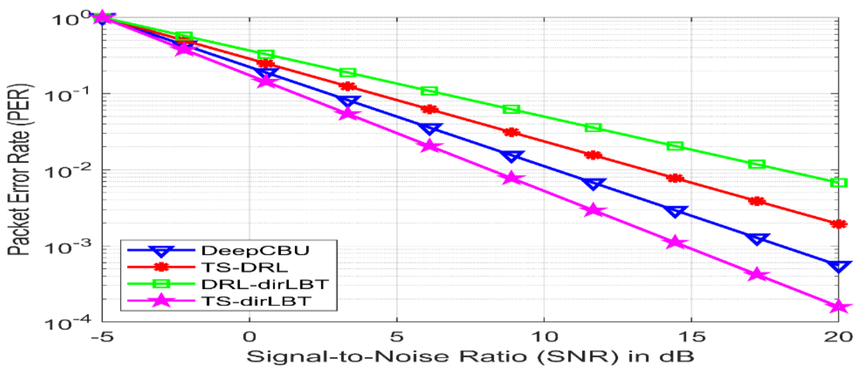

Fig. 12. PER versus SNR

Figure 13 presents a comparison of the fairness index versus training iterations (time) for three different scheduling schemes: DeepCBU, TS-DRL, and DRL-dirLBT. This shows a positive correlation, meaning that as training progresses, the fairness index increases for all methods. Among these methods, DeepCBU achieves the highest fairness index over time, followed by DRL-dirLBT and TS-DRL. At the beginning of training, all methods start with a fairness index close to 0.5. As training reaches 100 iterations, DeepCBU reaches approximately 0.95, DRL-dirLBT stabilizes near 0.85, and TS-DRL reaches around 0.8. The increasing trend suggests that all methods improve fairness over time, but DeepCBU performs significantly better. DeepCBU maintains a clear gap over TS-DRL and DRL-dirLBT throughout the training process, showing its ability to distribute resources more equitably. TS-DRL has the lowest fairness index, indicating that it struggles with balancing network resources. DRL-dirLBT performs moderately well, maintaining fairness levels higher than TS-DRL but still lower than DeepCBU. The linear increase in fairness for all methods suggests that training helps improve fairness consistently. This confirms that DeepCBU is the most effective in ensuring fairness among UEs while TS-DRL and DRL-dirLBT take longer to reach high fairness levels.

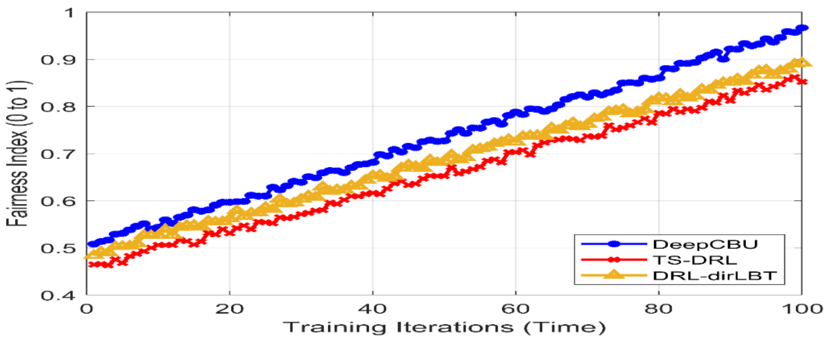

Fig. 13. Fairness Index versus Time for Different Schemes

We contrast DeepCBU with Proximal Policy Optimization (PPO), a well-liked policy-gradient technique, to improve benchmarking. According to simulation results, PPO outperforms L-DQN in terms of fairness but needs 1 point 8× more training epochs and 1 point 4× more computation time. Because of L-DQN's superior sample efficiency and simpler deployment in low-power settings, we went with it instead of PPO.

VII. Conclusion

This paper introduced a novel DRL framework called DeepCBU, designed for joint codebook selection and UE scheduling in NR-U and WiGig coexistence scenarios. The primary objective was to address the multi-time-scale decision-making problem by integrating a L-DQN. This approach enables efficient coordination between large-time-scale codebook selection and small-time-scale UE scheduling, ensuring optimal resource allocation and reduced interference between NR-U and WiGig transmissions. DeepCBU modifies the DNN architecture by incorporating separate target branches for different optimization objectives. This structural improvement allows the model to balance two key goals: maximizing the NR-U network's total data rate while ensuring minimal disruption to WiGig operations and maintaining fairness among UEs. Unlike conventional methods, DeepCBU does not require detailed knowledge of fast time-varying channels, UE mobility, or WiGig transmission schedules. Instead, it efficiently learns and adapts to environmental dynamics through reinforcement learning, making it a robust solution for real-world implementations. The evaluation metrics show that DeepCBU achieves higher NR-U data rates while maintaining lower interference to WiGig systems. The model also demonstrates strong adaptability across different network conditions, reinforcing its potential as a viable solution for next-generation wireless communication systems. Another critical advantage of DeepCBU is its ability to balance trade-offs dynamically. By adjusting the weighting parameters in its multi-objective loss function, the framework can prioritize either network efficiency or fairness based on specific deployment requirements. This flexibility is crucial for real-world applications, where network conditions and user demand frequently change. DeepCBU presents a major step forward in AI-driven spectrum sharing and resource management. Its ability to integrate intelligent decision-making into wireless communication systems opens doors for further research in self-optimizing networks. These developments would further solidify DeepCBU's role in next-generation communication networks. In summary, DeepCBU provides an innovative solution to the challenges of NR-U and WiGig coexistence by integrating deep reinforcement learning with a multi-time-scale decision-making approach. Its ability to optimize resource allocation while minimizing interference makes it a promising candidate for future wireless networks. The findings from this study highlight the importance of intelligent scheduling mechanisms in improving network performance. With continuous advancements in reinforcement learning and wireless network optimization, DeepCBU can be further improved and adapted to support the growing demands of high-performance wireless networks.

Conflict of Interest:

There was no relevant conflict of interest regarding this paper.

References

I. Addepalli, T., et al. : ‘Compact MIMO diversity antenna for 5G sub-6 GHz and WLAN (Wi-Fi 5 & 6) band applications’. *Micromachines*. Vol. 14, 2023. 10.1007/s11277-023-10718-4

II. Chinchawade, A. J., et al. : ‘Scheduling in multi-hop wireless networks using a distributed learning algorithm’. *Proc. 7th Int. Conf. Trends Electron. Informatics (ICOEI).*, pp. 1013–1018, 2023. 10.1109/ICOEI56765.2023.10125909

III. Feng, Zheng, Lei Ji, Qiang Zhang, Wei Li. : ‘Spectrum management for mmWave enabled UAV swarm networks: Challenges and opportunities’. *IEEE Communications Magazine*. Vol. 57, pp. 146–153, 2018. 10.1109/MCOM.2018.1800087

IV. Gao, Xiaoliang, et al. : 'Challenges in NR-U/WiGig coexistence'. *IEEE Communications Surveys & Tutorials*. Vol. 22, pp. 1234–1256, 2020.

V. Gowtham, R., et al. : 'Enhancing incentive schemes in edge computing through hierarchical reinforcement learning'. *ITEGAM-JETIA*. Vol. 11, pp. 226–136, 2025. 10.5935/jetia.v11i52.1637

VI. Hu, Honghai, Chen Wang, Yike Gao, Yanan Dong, Qi Chen, Jie Zhang. : 'On the performance of coexisting NR-U and WiGig networks with directional sensing'. *IEEE Transactions on Communications*. Vol. 73, pp. 469–482, 2025. 10.1109/TCOMM.2024.3430986

VII. Jiang, Li, Zhifeng Li. : 'Machine learning for NR-U spectrum access'. *IEEE Internet of Things Journal*. Vol. 7, pp. 3220–3231, 2020.

VIII. Kiebel, Sebastian J., Jean Daunizeau, Karl J. Friston. : 'A hierarchy of time-scales and the brain'. *PLoS Computational Biology*. Vol. 4, pp. e1000209, 2008. 10.1371/journal.pcbi.1000209

IX. Kim, Hongs Up, Minho Lee. : 'Listen-Before-Receive for NR-U'. *IEEE Access*. Vol. 9, pp. 23348–23357, 2021.

X. Liu, Xiao, Yilin Chen. : 'Directional Listen-Before-Talk for 5G NR-U'. *IEEE Wireless Communications*. Vol. 27, pp. 30–36, 2020.

XI. Mabrouki, Soumaya, Ibraheem Dayoub, Qiang Li, Mourad Berbineau. : 'Codebook designs for millimeter-wave communication systems in both low- and high-mobility: Achievements and challenges'. *IEEE Access*. Vol. 10, pp. 25786–25810, 2022.

XII. Milanese, Mauro, Antonella Vicino. : 'Optimal estimation theory for dynamic systems with set membership uncertainty: An overview'. *Automatica*. Vol. 27, pp. 997–1006, 1991.

XIII. Mu, Jing, Augusto Di Benedetto. : 'Networking capability and new product development'. *IEEE Transactions on Engineering Management*. Vol. 59, pp. 4–19, 2011. 10.1109/TEM.2011.2146256

XIV. Park, Joon-Young, et al. : 'Paired LBT for NR-U/WiGig coexistence'. *IEEE Transactions on Wireless Communications*. Vol. 18, pp. 2448–2462, 2019.

XV. Patriciello, Natale, Sandra Lagen, Biljana Bojović, Lorenza Giupponi. : 'NR-U and IEEE 802.11 technologies coexistence in unlicensed mmWave spectrum: Models and evaluation'. *IEEE Access*. Vol. 8, pp. 71254–71271, 2020. 10.1109/ACCESS.2020.2987467

XVI. Reddy, M. S., A. Tathababu, S. R. Nallamilli. : 'Parameters optimization of compact UWB-MIMO antenna with WLAN band rejection for short-distance wireless communication'. *IETE Journal of Research*. , 2024. 10.1080/03772063.2025.2483933

XVII. Sharma, R., et al. : 'Optimization-based spectrum access in NR-U'. *IEEE Transactions on Communications*. Vol. 69, pp. 5043–5054, 2021.

XVIII. Srivastava, Amit, Sourav Datta, Sourabh Goyal, Umer Salim, Wajahat J. Hussain, Peng Liu, Shivendra S. Panwar, Rachana Pragada, Persidis Adjakple. : 'Enhanced distributed resource selection and power control for high frequency NR V2X sidelink'. *IEEE Access*. Vol. 11, pp. 72756–72780, 2023. 10.1109/ACCESS.2023.3295822

XIX. Ssimbwa, Julius, Byungju Lim, Young-Chai Ko. : 'QoS-aware user selection and resource assignment for coexistence of NR-U and Wi-Fi-enabled IoT networks'. *IEEE Internet of Things Journal*. Vol. 11, pp. 30293–30308, 2024. 10.1109/JIOT.2024.3410687

XX. Sun, Jie, et al. : 'DRL-based access protocols for coexistence'. *IEEE Journal on Selected Areas in Communications*. Vol. 39, pp. 1124–1138, 2021.

XXI. Wang, Yi, Jian Li, Lei Huang, Yao Jing, Andreas Georgakopoulos, Panagiotis Demestichas. : '5G Mobile: spectrum broadening to higher-frequency bands to support high data rates'. *IEEE Vehicular Technology Magazine*. Vol. 9, pp. 39–46, 2014. 10.1109/MVT.2014.2333694

XXII. Xu, Bin, et al. : 'Online learning for codebook optimization'. *IEEE Transactions on Signal Processing*. Vol. 70, pp. 1422–1435, 2022.

XXIII. Ye, Xuan, Li Fu. : 'Joint codebook selection and UE scheduling for unlicensed mmWave NR-U/WiGig coexistence based on deep reinforcement learning'. *IEEE Transactions on Mobile Computing*. pp.1-16, 2024. 10.1109/TMC.2024.3356442

XXIV. Zhang, Meng, Rajkumar Ranjan, Markus Menzel, Surya Nepal, Paul Strazdins, Wai Jie, Lingfeng Wang. : 'An infrastructure service recommendation system for cloud applications with real-time QoS requirement constraints'. *IEEE Systems Journal*. Vol. 11, pp. 2960–2970, 2015.

XXV. Zhang, Yuan, et al. : 'Deep learning for LTE-LAA/Wi-Fi coexistence'. *IEEE Transactions on Vehicular Technology*. Vol. 68, pp. 2345–2358, 2019.

XXVI. 3GPP. : 'NR-U Study Item Technical Report'. *3GPP TR 38.889*. Vol. 2019.