IEEE COMPUTER SOCIETY

Dynamic Spectrum Coexistence of NR-V2X and Wi-Fi 6E using Deep Reinforcement Learning

Kashish D. Shah[1], Student Member, IEEE, Dhaval K. Patel[1], Senior Member, IEEE, Brijesh Soni[2], Member, IEEE, Siddhartan Govindasamy[3], Member, IEEE, Mehul S. Raval[1], Senior Member, IEEE, Mukesh Zaveri[4], Member, IEEE

[1]School of Engineering and Applied Science, Ahmedabad University, Navrangpura, Ahmedabad 380015, Gujarat, India
[2]Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43065, U.S.A
[3]Department of Engineering, Boston College, Newton, MA 02467, U.S.A
[4]Department of Computer Science and Engineering, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, Gujarat, India

CORRESPONDING AUTHOR: Kashish D. Shah (e-mail: kashish.s2@ahduni.edu.in).

ABSTRACT The deployment of 5G NR-based Cellular-V2X, i.e., the NR-V2X standard, is a promising solution to meet the increasing demand for vehicular data transmission in the low-frequency spectrum. The high throughput requirement of NR-V2X users can be overcome by extending it to utilize the sub-6 GHz unlicensed spectrum, coexisting with Wi-Fi 6E, thus increasing the overall spectrum availability. Most existing works on coexistence rely on rule-based approaches or classical machine learning algorithms. These approaches may fall short in real-time environments where adaptive decision-making is required. In this context, we introduce a novel Deep Reinforcement learning (DRL) based framework for 5G NR-V2X (mode-1 and mode-2) and Wi-Fi 6E coexistence. We propose an algorithm to dynamically adjust the transmission time of the 5G NR-V2X (for mode-1) or Wi-Fi 6E (for mode-2), based on the Wi-Fi and V2X traffic, to maximize the overall throughput of both systems. The proposed algorithm is implemented through extensive simulations using the Network Simulator-3 (ns-3), integrated with a custom Deep Reinforcement Learning (DRL) framework developed using OpenAIGym. This closed-loop integration enables realistic, dynamic interaction between the learning agent and high-fidelity network environments, representing a novel simulation setup for studying NR-V2X and Wi-Fi coexistence. The results show that when employing DRL on NR-V2X and Wi-Fi coexistence, the average data rates for Vehicular User Equipments (VUEs) and Wi-Fi User Equipments (WUEs) improve by $\sim$ 24% and 23%, respectively, as compared to the static method; and even higher improvement when compared to the existing RL-based LTE-V2X and Wi-Fi coexistence approach. Additionally, we analyzed the impact of NR-V2X coexistence on the Wi-Fi subsystem under mode-1 and mode-2 communications. Our findings indicate that mode-1 communication demands more spectrum resources than mode-2, leading to a performance compromise for Wi-Fi.

INDEX TERMS Deep Reinforcement Learning, C-V2X, 5G NR-V2X, Spectrum Coexistence, Wi-Fi 6E.

I. Introduction

Intelligent Transportation Systems (ITS) has emerged as a critical solution for enhancing vehicular safety and reducing accidents through the integration of advanced communication technologies and real-time decision-making. The Cellular-Vehicle to-everything (C-V2X) systems play a significant role in supporting vehicular communication in ITS. C-V2X encompasses two key cellular vehicular standards [1] i.e., LTE-V2X and 5G NR-V2X.

A key challenge faced by C-V2X systems lies in facilitating high-data transmission amid growing vehicular density and escalating vehicular data volumes, including video and LIDAR data.

Moreover, in 2020, the Federal Communication Commission (FCC) changed the ITS service spectrum band allocation from 5.855 - 5.925 GHz to 5.895 − 5.925 GHz, i.e., reducing the spectrum from 70 MHz to 30 MHz [2]. Consequently, meeting the high spectrum requirements of C-V2X has become challenging. The FCC [2] in the US and regulators in Europe [3] allowed unlicensed operations in the 6 GHz band. Wi-Fi, being one of the key unlicensed technologies, is entering the 6 GHz band as per the IEEE 802.11 Working Group [4]. The Wi-Fi 6, i.e., 802.11ax in the 6 GHZ band, is also stated as Wi-Fi 6E. The C-V2X technology can potentially coexist with Wi-Fi 6E in the sub-6GHz spectrum to satisfy spectrum needs.

A. Current state of the art and Motivation

Several works have addressed the coexistence of Wi-Fi with LTE-based technology. For instance, an adaptive listen-before-talk (LBT) based algorithm was proposed in [5], requiring LTE-U to know the channel at the subframe edge and select a new idle channel. A fair LBT algorithm was also suggested in [6], which assigns an appropriate idle

period to Wi-Fi to ensure transmission and integrate system throughput with fairness between LTE-U and Wi-Fi. The LTE-U forum proposed the carrier sensing and adaptive duty cycle-based transmission algorithm (CAST) [7], where the small cell performs channel sensing on all available unlicensed channels and selects the idlest channel based on media activity observations.

However, LTE-Wi-Fi coexistence mechanisms fall short of meeting the stringent requirements of next-generation vehicular networks. Hence, the 5G NR-based Cellular-V2X standard was advanced in 3GPP Rel. 16 [8].

In 2021, Naik et. al. in [9] proposed the study of modeling the impact of the multi-user OFDMA feature introduced in 802.11ax on coexisting with 5G NR-U. Similarly, a framework to find the optimal fairness parameters for 5G-NR U, coexisting with Wi-Fi, was proposed in [10]. Later, in the context of V2X, a decentralized coexistence protocol for V2X and Wi-fi was proposed in [11]. Additionally, mitigation techniques for co-channel co-existence of 802.11p with NR-V2X SideLink (SL) mode were studied in [12]. Yet, these are static techniques and cannot update the channel access of either of the systems dynamically.

With the rise of learning-based signal processing, data-driven methods have attracted significant interest for C-V2X (LTE-V2X, NR-V2X) and Wi-Fi coexistence scenarios [13]. Leveraging their strong learning capabilities, several studies have applied Machine Learning (ML) to address coexistence challenges. For example, [14] proposed a Q-learning-based duty cycle adaptation algorithm for a single unlicensed channel, achieving optimal convergence. Similarly, [15] presented a Q-learning approach for LTE-U and Wi-Fi coexistence in multi-channel settings. Authors in [16] developed a duty cycle-based adaptive algorithm using Reinforcement Learning (RL) to dynamically select the duty cycle of the co-channels used by both the LTE-V2X and Wi-Fi systems. Here, the authors considered a discrete state and action space for selecting duty cycles. Often, the information state space is continuous. Basic value or policy iteration cannot be applied to a continuous state space, and thus, Deep Reinforcement Learning (DRL) can be employed. For instance, authors in [17] propose a DRL-based coexistence scheme for LAA-LTE and Wi-Fi. Similarly, Pei et al. in [18] proposed a mean field-based DRL approach for LTE-U and Wi-Fi coexistence, adding a game-theoretic aspect.

Deploying 5G NR-V2X in the sub-6GHz unlicensed spectrum presents critical challenges, specifically due to interference between Vehicular User Equipment (VUEs) and Wi-Fi 6E users (WUEs). The coexistence issue is intensified by both systems' fundamentally different channel access mechanisms, leading to frequent contention and unpredictable performance degradation. Wi-Fi employs Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), which relies on contention-based access. At the same time, NR-V2X utilizes Dynamic Grant (DG) and Configured Grant (CG) scheduling in mode-1 or Semi-Persistent Scheduling (SPS) in mode-2 [1]. These differences increase the likelihood of packet collisions, as both NR-V2X UEs and Wi-Fi STAs may attempt to access the same time-frequency resources without coordination. The absence of a unified spectrum-sharing strategy leads to QoS degradation in high-mobility vehicular scenarios, where latency and reliability are critical. NR-V2X may face packet loss, while Wi-Fi experiences reduced throughput. A potential solution is a high-level frame-based MAC protocol to regulate spectrum access and minimize collisions, but designing such a framework while meeting both systems' performance needs remains a challenge.

Traditional spectrum management relies on static, rule-based methods or classical ML models, which are sub-optimal in dynamic, real-time environments. These approaches struggle in high-traffic scenarios with fluctuating network conditions. Reinforcement Learning (RL) offers a more adaptive solution by making sequential decisions to optimize resource allocation. This work leverages RL to address the challenges of dynamic spectrum coexistence.

In contrast to the existing approaches, such as the Q-learning-based methods in [15, 16] or the DRL-based strategies with simplified MAC designs in [17, 18], our work addresses the coexistence problem using a more expressive and adaptive learning framework. These prior works typically operate with discrete state-action spaces or static duty-cycle tuning, which limits their responsiveness in real-time, high-mobility environments, particularly in vehicular networks where traffic and interference conditions vary rapidly. Moreover, they do not adequately support continuous state observations or flexible spectrum access. Thus, we model coexistence as a sequential decision-making problem, where actions are selected based on real-time, *continuous-valued states*. Continuous-valued states are critical to accurately capture the performance of both NR-V2X and Wi-Fi systems. This formulation leads to a computationally hard problem, which we address using Deep Reinforcement Learning (DRL) over a Markov Decision Process (MDP). Our DRL framework enables fine-grained, adaptive control of transmission parameters, overcoming the limitations of static and discrete models while ensuring robust coexistence in realistic network conditions. The full framework is detailed in Section III.

To the best of the authors' knowledge, such an approach, which considers DRL for NR-V2X and Wi-Fi 6E coexistence under the sub-6 GHz unlicensed spectrum, although promising, has not been reported in the literature. In this context, we propose a novel DRL-based coexistence framework to dynamically allocate resources between NR-V2X and Wi-Fi by conducting network simulations in ns-3.

B. Contributions

The key contributions of this work are fourfold and can be summarized as:

- We propose a novel frame-based MAC protocol for the NR-V2X and Wi-Fi coexistence. This protocol

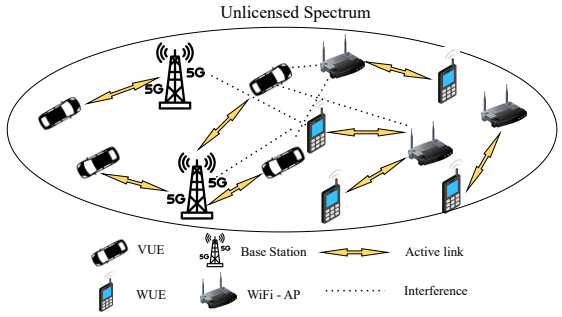


FIGURE 1: 5G NR-V2X mode 1 and Wi-Fi coexistence network model.

is proposed on top of the regular MAC interface of the NR-V2X (both modes) and Wi-Fi. It opens an advantage of getting control of the transmission time of the systems, which turns it into a decision-making problem.

- To address the coexistence between NR-V2X and Wi-Fi, considering a more informative continuous state space, and enabling dynamic decision making, we propose an innovative DRL-based technique (along with a novel distance-based reward) for allocating the transmission time to the V2X and Wi-Fi networks. The proposed algorithm helps achieve a higher cumulative reward, improving both systems' overall throughput and fair coexistence.
- We develop a custom integrated framework by integrating ns-3 version - 3.40, with a programmed DRL algorithm to enable closed-loop and online interaction between network simulation and DRL-based decision-making. This integration allows the DRL agent to learn from realistic NR-V2X and Wi-Fi coexistence scenarios, providing a practical and flexible environment for training and evaluation.
- We compare the proposed approach with the LTE-V2X and analyze the impact of coexistence on both the sub-systems. First, the 5G NR-V2X standard is compared with the LTE-V2X standard when modes 1 and 2 coexist with Wi-Fi. Secondly, we analyzed the impact on the Wi-Fi subsystem when coexisting with NR-V2X mode-1 and mode-2 systems. Numerical results state that when employing DRL on NR-V2X and Wi-Fi coexistence, the average data rates for VUEs and WUEs improve by $\sim$ 24% and 23%, respectively, as compared to the static method (defined later); and even higher improvement when compared to the existing RL-based LTE-V2X and Wi-Fi coexistence approach.

II. System Model and Problem Formulation

A. NR-V2X mode-1 and Wi-Fi Coexistence System

We consider the system model shown in Figure 1. The VUEs are the NR-V2X users, and the WUEs are the Wi-Fi Users (Wi-Fi stations). Both kinds of users cooperate and share the unlicensed spectrum. We consider the system to have a single Wi-Fi access point and gNB.

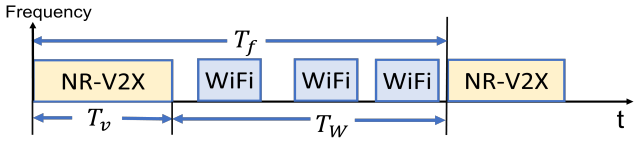


FIGURE 2: The frame-based MAC protocol for the coexistence system.

The NR-V2X system follows the 5G V2X communication with the 5G NR air interface, whereas the Wi-Fi subsystem will use the standard CSMA/CA mechanism. To support the spectrum coexistence between these two technologies, we propose the frame-based MAC protocol as shown in Figure 2. This MAC protocol is a higher-level protocol to prevent simultaneous access by the two technologies. The Wi-Fi's Transmit Opportunity (TxOP) or NR-V2X's centralized scheduling (for mode-1) or Semi-persistence Scheduling (SPS) (for mode-2) are not modified. Thus, there can be multiple TxOPS for Wi-Fi within a defined Wi-Fi transmission period 'T_w'. The NR-V2X users transmit for the first T_v interval by taking access to the channel at the start of the current frame to ensure that no Wi-Fi transmissions occur that overlap. This can be ensured by making the gNB immediately schedule (without sensing) at the start of the frame, scheduling the VUEs for uplink transmissions, and sending the resource allocation information while the Wi-Fi is still sensing. Furthermore, to ensure that no WUE transmission starts towards the end of the T_w duration, which will spill over to the start of the next frame, the NR-V2X can send a control frame near the end of the frame time. This control frame will probably experience a collision, but with that, Wi-Fi will perform backoff as per CSMA/CA and thus would not interfere at the start of the next frame. During the transmission in the NR-V2X subsystem, the channel will be completely utilized, and the WUEs will perform a standard back-off and wait for the channel to get free. Once the channel is empty when the transmission of the NR-V2X subsystem is ended, the WUEs (Wi-Fi stations) will start contending as per the CSMA/CA mechanism for the next T_w time interval. The total frame time is T_f. Thus,

$$T_f = T_v + T_w \tag{1}$$

B. Problem Formulation

We address the following two cases with this new approach. If the transmission time $T_v \gg T_w$ even though there is less VUE traffic, the Wi-Fi subsystem may suffer from packet loss and collisions in case of high WUE traffic. On the other side, if we keep $T_w \gg T_v$, the aggregate throughput of the NR-V2X system will be affected and reduced significantly. Since the NR-V2X subsystem may be running safety-critical applications, this communication loss may be detrimental. Thus, by learning the traffic pattern on both the subsystems and considering that the system is deployed on the NR-V2X gNB, an approach to dynamically select the T_v is needed. Let us assume the R_v and R_w are the aggregated throughput of the NR-V2X and Wi-Fi systems, respectively. The problem

can be formulated as follows:

$$\max_{T_v} R_v + R_w$$
$$s.t.\ T_v + T_w = T_f \quad (2)$$

The total frame time T_f in the above equation is fixed. Now, we assume the system is deployed on the base station. Thus, getting the aggregated Wi-Fi throughput at the base station is challenging. The aggregated throughput of the NR-V2X subsystem can be maintained by analyzing the total packets transmitted and received within the given time interval. But as mentioned in [17], the information to track the Wi-Fi traffic pattern, like the number of collisions and successful transmissions, can be extracted by applying advanced energy detection techniques on the channel. Thus, we use the NR-V2X throughput and Wi-Fi features extracted from the channel energy management to create a DRL approach that dynamically slides over the time window T_v to maintain fair coexistence between NR-V2X and Wi-Fi as discussed in the next section.

III. Proposed Deep-RL based NR-V2X and Wi-Fi Coexistence

This section presents the overall process to solve the above-formulated problem, which can be regarded as a Markov Decision Process (MDP). Later in the following subsections, we propose a Deep Q-Learning-based algorithm for the co-existence of NR-V2X and Wi-Fi.

A. Defining the Markov Decision Process

We develop a system at the gNB that can control the transmission time of the NR-V2X subsystem within a frame duration. This system has access to the available features of the NR-V2X and Wi-Fi, which can be analyzed to identify the traffic pattern of both the subsystems working in the same unlicensed spectrum and make a decision on choosing the transmission time T_v of the NR-V2X subsystem. The observation state, i.e., the features available to gNB and choosing T_v at a particular time, are independent of the previous state and actions. Thus, this process can be well addressed as an MDP, with the gNB as the decision maker.

MDP is a decision-making framework that models stochastic environments. It comprises a tuple of five elements (defined in the upcoming sections), including state space S, action space A, transition probability $P_{a_t}(s_t, s_{t+1})$, reward $R_t(s_t, s_{t+1})$, and discount factor γ. The decision-maker, known as the agent in MDP, observes the environment's states, which comprise state space S. The action space A contains all the agent's available actions. Transition probability $P_{a_t}(s_t, s_{t+1})$ determines the possibility that the agent's action at the state st will lead to the state s_{t+1}. Immediate reward $R_t(s_t, s_{t+1})$ is the reward the agent receives from transitioning from s_t to s_{t+1} by taking action a_t. The discount factor $\gamma \in [0, 1)$ represents the extent to which future rewards are considered in present decisions. The primary objective of MDP is to find the optimal policy $\pi^\star(s)$ that maximizes the long-term cumulative reward $\sum_{t=0}^{\infty} \gamma \times R_t(s_t, s_{t+1})$. The solution of MDP, i.e., $\pi^\star(s)$, can be obtained through dynamic programming (DP) methods, such as value iteration, which require complete knowledge of the system dynamics, including the transition probability $P_{a_t}(s_t, s_{t+1})$. DP methods are called model-based RL techniques, while those that can solve MDP without knowing the system dynamics are called model-free RL techniques. Considering the problem formulated in (2), it is impossible to apply Q-learning techniques as the state space is not discrete but continuous.

B. DRL-based Coexistence Algorithm for NR-V2X mode-1

The proposed framework for NR-V2X and Wi-Fi coexistence is illustrated in Figure 3. At any time step 't', the Deep-Q Reinforcement Learning (DQN) agent selects an action 'a_t' from the predefined action space 'A', determining the transmission time parameters T_v for NR-V2X and T_w for Wi-Fi 6E. These transmission parameters are then applied to ns-3 simulations, which model the interactions between the two subsystems under the chosen settings. The outcome of the ns-3 simulations is captured as the current observation or state 's_t', representing the system's status at time 't'. The RL agent predicts the optimal transmission timing using the learned policy network based on this state. Alternatively, it may select a random action as part of the exploration-exploitation trade-off. This newly chosen action is again fed into the ns-3 simulation to generate the next state 's_{t+1}'. A reward 'r_{t+1}' is then computed based on the transition from 's_t' to 's_{t+1}', assessing how well the selected action improves system performance. This transition, consisting of 's_t', 'a_t', 'r_{t+1}', and 's_{t+1}', forms an experience tuple 'e_t', which is stored in the replay memory. The agent leverages this replay memory to refine its decision-making policy, aiming to maximize the cumulative discounted reward over time. In the following section, we define the state space, action space, and reward function, which are crucial for formulating the problem as described in (2). Considering the gNB as an agent, we model NR-V2X and Wi-Fi networks as stochastic environments. The target of the agent is to dynamically manage the transmission time of the NR-V2X T_v system in every frame. Based on that, the Wi-Fi transmission time will be decided per (1), as the total frame time T_f is fixed. Therefore, the agent will increase or decrease the T_v after learning the traffic on both subsystems from the observation/state space.

$$A = \{-50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 50\} \quad (3)$$

The action space in (3) (also in Figure 3) is the required change in T_v in milliseconds (ms). For example, the action of '-50' refers to the reduction of T_v by 50ms. This implies that there will be an increase of 50ms in T_w. This action should ideally be taken in cases with more traffic on the Wi-Fi than on the NR-V2X side. We need the agent to learn an optimal policy to take good actions. By looking at the observations/states and taking actions randomly (initially),

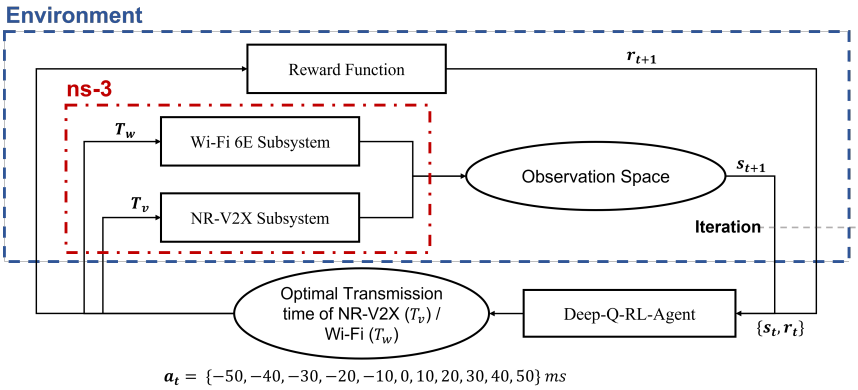


FIGURE 3: The proposed framework for the NR-V2X and Wi-Fi coexistence system.

the agent can learn the Q-values for every state-action pair. Therefore, it is necessary to define an effective state space for the agent. Now, as discussed in the earlier sections, not considering the aggregate throughput of the Wi-Fi system is a more realistic scenario. However, we need to identify key features that will enable the agent to effectively gather information about WUE traffic. Accordingly, we define the following state or observation space from Figure 3:

$$s = \{n_c, n_t, R_v\} \tag{4}$$

where n_c, n_t, and R_v are the number of collisions during Wi-Fi transmissions, successful Wi-Fi transmissions, and the aggregate throughput of NR-V2X users. Here's a refined version of the sentence for improved clarity:

These parameters can be obtained using advanced energy detection techniques applied to the channel during Wi-Fi transmissions, as outlined in [17]. In our case, we extracted them from the channel information during the simulation. As the gNB itself is the agent, and that's where we deploy the framework, R_v can be calculated by analyzing the total packets transmitted and received within the given time interval during the NR-V2X transmission.

For every state, the agent will choose an action. As in most RL systems, the agent will initially choose actions randomly. However, with increasing iterations, the agent learns the appropriate action to take in each specific state, which is known as a policy. To develop an effective policy, the agent requires feedback after every action. This feedback can be given by providing the agent with a reward. As per the features extracted in the state space, we propose the following reward function to make the agent learn the optimal policy by maximizing the cumulative discounted reward.

Our key innovation lies in the distance-based reward function (Eqs. 5–9), which uniquely optimizes fairness by minimizing Manhattan distances between current and ideal states (n_c, n_t, R_v). This approach outperforms traditional throughput-maximization rewards by penalizing collisions and prioritizing balanced resource allocation. The decentralized DRL architecture for mode-2 (Figure 4) also eliminates reliance on gNB, enabling AP-driven decisions—a first in NR-V2X/Wi-Fi coexistence literature. The Manhattan distance will be calculated between the individual state variables: n_c, n_t, R_v, and other max/min variables $n_{c,min}$, $n_{t,max}$, $R_{v,max}$. With the distance-based reward, the agent will get a greater reward only if the state space variables n_c, n_t, and R_v are closer to the max/min values reached till that point. To initiate learning, a higher value will be initialized for $n_{t,max}$ and $R_{v,max}$, and a lower value for $n_{c,min}$. Consequently, the agent will be trained to select actions that maximize rewards, leading to more successful transmissions and fewer collisions, thereby improving the overall data rates of both systems. We define the distance-based functions as follows:

$$D_{n_c} = \left\{ \begin{array}{ll} 0, & \text{if } n_{c,min} > n_c \\ \mathcal{M}(n_c, n_{c,min}), & \text{otherwise} \end{array} \right\} \tag{5}$$

$$D_{n_s} = \left\{ \begin{array}{ll} 0, & \text{if } n_{t,max} < n_t \\ \mathcal{M}(n_t, n_{t,max}), & \text{otherwise} \end{array} \right\} \tag{6}$$

$$D_{R_v} = \left\{ \begin{array}{ll} 0, & \text{if } R_{v,max} < R_v \\ \mathcal{M}(R_v, R_{v,max}), & \text{otherwise} \end{array} \right\} \tag{7}$$

where D_{n_c}, D_{n_s}, and D_{R_v} are the required Manhattan distances for collisions, successful transmissions, and V2X data rate. $\mathcal{M}$ is the Manhattan distance function. The overall distance with respect to the above-computed distances can be written as:

$$D_t = \sqrt{D_{R_v}^2 + D_{n_s}^2 + D_{n_c}^2} \tag{8}$$

We create a variable D_{max} to normalize the distance to use it in the reward function. Finally, we compute the reward function as follows:

$$R_t = e^{-\frac{min(D_t, D_{max})}{D_{max}}} \tag{9}$$

The $D_{\max}$ variable will be updated to D_t if it is exceeded by D_t. Within the context of DRL, a Deep Neural Network (DNN), also known as Deep Q-Network (DQN), is created to estimate the expected cumulative reward of taking an action (a) in a particular state (s), denoted as $Q(s, a; \theta)$, where θ refers to the weights of the DQN. Once the DQN has converged with θ at $\theta^\star$, it can predict the maximum expected cumulative reward for any given state-action pair s, a as

$Q(s, a; \theta^\star)$. The optimal policy $\pi^\star(s)$ can be determined using this information.

$$\pi^\star = arg\max_a Q^\star(s, a; \theta^\star) \tag{10}$$

The typical DRL approach is based on the Q-learning framework [19], a model-free reinforcement learning algorithm. Specifically, we aim to learn the Q-function, Q_θ, parameterized by θ, which estimates the expected cumulative reward for taking an action a in a given state sequence $s_{\leq t}$ and following the optimal policy thereafter. The expected cumulative reward is the feedback to the agent for its selected transmission time T_v (or T_w for NR-V2X mode-2) by observing the current Wi-Fi and V2X traffic.
The Q-function is formally defined as [20]:

$$Q_\theta\left(s_{\leq t}, a\right) = \mathbb{E}\left[\sum_{\tau=0}^{\infty} \gamma^\tau r_{t+\tau} \mid S_0 = s_0, \ldots, S_t = s_t, A_t = a_t\right] \tag{11}$$

where r_t is the reward received at time step t, and γ is the discount factor that determines the importance of future rewards. The Q-function captures the expected sum of discounted rewards starting from the state sequence $s_{\leq t}$, taking action a, and thereafter following the optimal policy.

To update the Q-function a Bellman operator $\mathcal{B}$ is used, which is defined as:

$$\mathcal{B}Q_\theta\left(s_{\leq t}, a_t\right) = r_t + \gamma \max_a Q_\theta\left(s_{\leq t+1}, a\right) \tag{12}$$

The Bellman operator provides a recursive relationship for the Q-function, expressing the value of the current state-action pair in terms of the immediate reward plus the discounted maximum future reward.

An optimal policy π_L^* satisfies the Bellman equation:

$$\mathcal{B}Q_\theta^{\pi_L^*}\left(s_{\leq t}, a_t\right) = Q_\theta^{\pi_L^*}\left(s_{\leq t}, a_t\right) \tag{13}$$

where $Q_\theta^{\pi_L^*}$ is the Q-function corresponding to the optimal policy. We denote the optimal Q-function as $Q^* = Q^{\pi_L^*}$ when this condition is met. The optimal Q-function mentioned in the above equation by $Q_\theta^{\pi_L^*}$ can be learned from the following equation:

$$\begin{aligned} Q(s_{t+1}, a_{t+1}) =& Q(s_t, a_t) + \alpha \times [R_{t+1} \\ &+ \gamma(Q(s_{t+1}, a) - Q(s_t, a_t))] \end{aligned} \tag{14}$$

where α is the learning rate, controlling how much we adjust the network's weights during training, and γ is the discount factor, which determines how much importance we give to future rewards compared to immediate rewards (the recent rewards are based on the latest WUE/VUE traffic). In our DRL framework, we train two Deep Q-networks—a policy network and a target network—to learn the optimal action-selection strategy. The policy network takes the current state (i.e., the Wi-Fi and NR-V2X network variables defined in (4)) as input and computes the Q-values for each possible action, resulting in a set $\{q(s, a_1), q(s, a_2), \ldots, q(s, a_n)\}$ for a total of n actions (in our case, $n = 11$) as defined in (3). These Q-values represent the expected cumulative reward for taking each action from the current state. B ased on these computed values, the agent selects the best action, or sometimes a random action, during the early training phase to encourage sufficient exploration. Once an action is taken, the environment provides the next state s_{t+1} along with a corresponding reward. The target network then processes the next state s_{t+1} and computes the desired Q-values, denoted as $Q(s_{t+1}, a)$. These values are a stable reference or target when updating the policy network. The target network contributes to training stability by ensuring that the target Q-values used in the loss calculation remain consistent over multiple iterations. Each experience, comprising the current state, the selected action, the reward, and the next state, is stored in a replay memory as a tuple

$$e_t =< s_t, a_t, R_{a_t}(s_t, s_{t+1}), s_{t+1} > \tag{15}$$

Over several epochs, these experiences are collected in a memory list M. Later, random mini batches drawn from this replay memory are used to update the policy network. This experience replay mechanism helps break the correlation between sequential data, leading to more stable learning. The policy network is updated by minimizing the loss between its predicted Q-values and the target Q-values computed by the target network. Over many iterations, this training

Algorithm 1 Proposed DRL Algorithm

Require: Training the DQN agent with the NR-V2X and Wi-Fi simulations in ns-3. The target is to learn the optimal weights on both the policy and the target networks. Both the Neural Networks' weights will be initialized randomly.

while $t > t_{\max}$ **do**
  if rand() $<$ **E then**
    Random action $a_t \in A$
  else
    $a_t \leftarrow \text{argmax}_a Q(s_t, a; \theta)$
  end if
  $T_v \leftarrow a_t$ ▷ for mode - 1
  $T_w \leftarrow T_f - T_v$ ▷ for mode - 1
  $T_w \leftarrow a_t$ ▷ for mode - 2
  $T_v \leftarrow T_f - T_w$ ▷ for mode - 2
  $VUEs \leftarrow \text{Poisson}(\lambda)$
  $WUEs \leftarrow \text{Poisson}(\lambda)$
  $R_v \leftarrow \text{RunNR-V2XSimu}(T_v, VUEs)$
  $n_c, n_t \leftarrow \text{RunWi-FiSim}(T_w, WUEs)$
  $s_{t+1} \leftarrow n_c, n_t, R_v$ ▷ for mode - 2
  $s_{t+1} \leftarrow n_c, n_t$ ▷ for mode - 2
  $e_t \leftarrow \langle s_t, a_t, R_{a_t}(s_t, s_{t+1}), s_{t+1}\rangle$
  if $t \geq m$ **then**
    Get the e_t by getting m experiences from the replay memory list M.
    Update θ by minimizing the loss of the policy network based on the Q-function using SGD.
  end if
  if $t\%K == 0$ **then**
    $\theta' = \theta$
  end if
  $t = t + 1$
  Compute ϵ_{t+1} as per (18)
end while

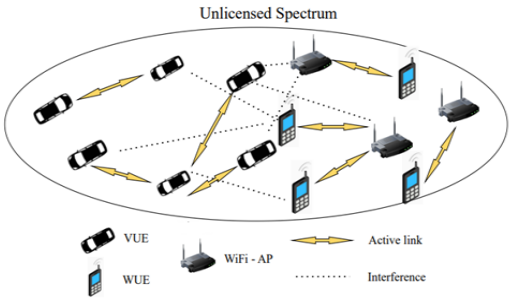


FIGURE 4: 5G NR-V2X mode-2 and Wi-Fi coexistence network model.

process enables the Q-function to converge toward the optimal policy. Initially, the agent selects actions randomly (exploration). Still, as training progresses and the epsilon value in the epsilon-greedy strategy decreases, the agent increasingly relies on the learned Q-values (exploitation). Periodic updates to the target network are performed by copying the weights from the policy network every K iterations. This further helps maintain training stability and avoid rapid fluctuations in the Q-value estimates. The training process is a continuous cycle in which the policy network estimates Q-values, actions are chosen and executed, new experiences are stored, and both networks are updated using these experiences until the Q-function converges to an optimal policy. The proposed DRL approach is described in Algorithm 1. It leverages a DQN agent, which undergoes training using simulations in ns-3 for both NR-V2X and Wi-Fi. The algorithm operates iteratively, with each iteration aiming to fine-tune the agent's decision-making process. At each time step t, the agent decides on an action a_t, which could either be a random in case the randomly generated value is less than a given probability $\mathbf{E} \in (0, 1)$ (line 2 in Algorithm - 1), or the action that maximizes the Q-value function $Q(s_t, a; \theta)$ for the current state s_t (line 5). The chosen action determines the transmission time T_v for the NR-V2X system (line 7, or line 10 for mode-2), with the remaining time allocated to the Wi-Fi system as T_w (line 8, or line 9 for mode-2).

The algorithm then simulates the environments for both NR-V2X and Wi-Fi. The NR-V2X simulation is executed with the transmission time T_v and VUEs arrival following a Poisson distribution with rate λ (line 11). Simultaneously, the Wi-Fi simulation runs with the remaining time T_w, and the WUEs arrival also follows a Poisson distribution with the same rate λ (line 12). The outcomes of these simulations, including the rewards and the new state s_{t+1}, are then used to update the agent's experience.

The agent collects experiences as tuples $\langle s_t, a_t, R_{a_t}(s_t, s_{t+1}), s_{t+1} \rangle$. Once enough experiences are gathered (i.e., $t \geq m$), it samples from the replay memory to update the policy network weights θ by minimizing the Q function loss using Stochastic Gradient Descent (SGD) (lines 19 and 20). At regular intervals K, the target network weights θ' are updated to match the policy network weights θ (line 20).

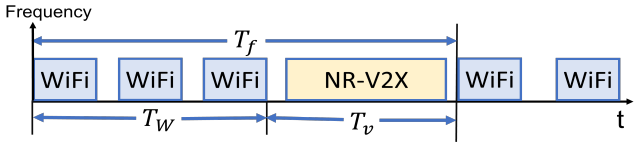


FIGURE 5: The frame structure of the coexistence system with NR-V2X mode - 2 SL

C. DRL-based Coexistence Algorithm for NR-V2X mode-2

The NR-V2X mode-2 SL does not use network coverage and is specifically designed for V2V and V2I SL communication. Thus, there is no involvement of the base station (gNB), and the vehicles and infrastructures having the V2X radio communicate with each other in the 5.9 GHz band using SPS scheduling. When modeling the NR-V2X mode 2 and Wi-Fi using RL, the question that needs to be addressed is what entity will play the role of an agent. The V2X network model without a gNB contains multiple VUEs communicating with each other via the V2V SL. The Wi-Fi, on the other hand, is the same as the one described in the earlier sections of this paper, employing the CSMA/CA as the channel access mechanism. The updated network model for the NR-V2X mode 2 and Wi-Fi coexistence is described in Figure 4.

The proposed frame-based channel access MAC protocol for mode-1 needs to be updated to solve the coexistence problem described in Figure 4, where there is no gNB. In the absence of the gNB on the V2X side communication, the control to execute the proposed frame-based MAC layer protocol is given to the Wi-Fi AP.

Figure 5 describes the updated frame-based protocol in which Wi-Fi will make the initial transmission for T_w time. This T_w time interval is decided by the RL Agent, i.e., the AP, as per the intelligence it receives in the state space. The problem formulation stays the same as (2), but the execution changes as the AP has the role of decision-making. We keep the same action space as the one described in (3), but the state space is updated as follows:

$$S = \{n_t, n_c\} \tag{16}$$

The NR-V2X Mode 2 is the offline mode of vehicular communication that will be extensively used for traffic safety and management purposes. Thus, it is necessary to ensure seamless transmission for this communication mode. For this reason, we propose a different frame-based channel access approach. We constrain the transmission time for the VUEs, stating that they will transmit for a fixed amount of time irrespective of the Wi-Fi and NR-V2X users' traffic. For a time frame T_f, the VUEs will transmit for at least $T_f/2$ units. If the Wi-Fi traffic is high, the Wi-Fi users will use other $T_f/2$ units; if the traffic is low, the Wi-Fi users will use $T_w < T_f/2$. Thus, the Wi-Fi AP, being the DRL agent, will choose the change in transmission time of Wi-Fi ΔT_w from the range $[0, T_w/2]$, by observing the state space mentioned above. The reward function will not depend on the NR-V2X parameters, as we are dedicating 50% of the transmission to the V2X system. We use the following reward function for

this modeling, which gives feedback to the agent based on the Wi-Fi transmission parameters:

$$R_t = e^{-\frac{n_t * n_c}{n_{t_{max}}}} \tag{17}$$

IV. Simulation setup and Results

In this section, we describe the ns-3 simulations of Wi-Fi and NR-V2X in Subsections 5-A and 5-B, respectively, DRL integration using the OpenAIGym framework in Subsection 5-C, and later describe the simulation results in the subsequent subsection. The Wi-Fi system consists of 1 AP and multiple WUEs, and the V2X subsystem consists of 1 gNB and multiple VUEs in case of mode-1, and no gNB with only VUEs in mode-2. The simulations are carried out per the Figure 3.

The ns-3 offers a dynamic environment for simulating diverse network scenarios, enabling the testing of newly proposed algorithms, protocols, frameworks, and models in a controlled yet realistic setting. We simulate the 5G NR mode-1 and mode-2 scenarios and the Wi-Fi 802.11 ax network using the available NR and Wi-Fi modules on the central frequency of 5.9 GHz.

We implement Algorithm 1 in the simulations, which require online learning of the Deep-RL agent. In ns-3, Wi-Fi simulations for the 802.11ax standard can be configured to operate in the 6 GHz frequency band. For NR-V2X simulations, the central frequency can be adjusted to accommodate custom frequencies. To study coexistence, NR-V2X simulations are set to run at 5.9 GHz with a bandwidth of 50 MHz. Table 1 outlines the simulation parameters for both the Wi-Fi and NR-V2X systems.

A. Simulation of Wi-Fi 802.11ax

We simulate the Wi-Fi Standard 802.11ax with multiple STAs and a single AP. This is the Wi-Fi subsystem of the framework described in Figure 3. We use the Wi-Fi simulations validated in [21]. This system is simulated based on the CSMA/CA with the number of incoming WUEs also following a Poisson process with the arrival rate of λ[1] For 802.11ax, ns-3 supports three types of channels with different widths: 20MHz, 40MHz, and 80 MHz. Thus, we consider all three for testing. We consider the Data and Control modes as 'OfdmRate54Mbps'.

B. Parameters of 5G NR-V2X Mode - 1 and Mode - 2 Simulations

To simulate the network coverage scenario characteristic of NR-V2X mode-1, we employ the nr-module proposed under the 5G-lena-v2x simulation by CTTC Spain [22]. The NR-V2X system is implemented on the 5G NR air interface utilizing the ns-3 nr-module. Our simulation scenario includes a single gNB and multiple VUEs, with the VUEs following a Poisson arrival process characterized by the arrival rate λ. This simulation contains multiple vehicular devices in three lanes, communicating via the V2V SL. The SPS technique is implemented as the MAC protocol of these devices, and on top of that, the proposed upper layer frame-based MAC protocol is implemented to get the transmission time T_v for the VUEs.

[1]Random traffic of VUEs and WUEs is generated based on the arrival rate λ. Equal arrival rates are considered in the simulations.

TABLE 1: Simulation parameters for NR-V2X and Wi-Fi 802.11ax.

Parameter	NR-V2X	Wi-Fi 802.11ax
Time for a Successful Tx: T_s	50 ms	50 ms
Total Frame Time: T_f	100×T_s	100×T_s
Arrival rate of Users	$\lambda = \{5.5, 6.5, \ldots, 9.5\}$	$\lambda = \{5.5, 6.5, \ldots, 9.5\}$
Number of gNBs	1	1
Modulation Scheme Index (MCS)	7	7
Simulation Time	T_v	T_w
Total Transmit Power	8 dBm	8 dBm
Bandwidth/Channel Width (CW)	50 MHz	20 MHz
Central Frequency	5.9 GHz	6 GHz
Packet Size	1252 bytes	1472 bytes

C. Deep-RL Integration

Once the Wi-FI and NR-V2X subsystems, as detailed in Section 5-A and Section 5-B, respectively, are configured using the ns-3 (ver 3.40) nr-module and OpenAIGym, they are invoked during the execution of the DRL setup, as illustrated in Figure 3 and Algorithm-1. We are the first to integrate ns-3 simulations with DRL specifically to develop a high-level MAC protocol. No other integration module is available for this specific purpose. The Wi-Fi and NR-V2X parameters n_c, n_t, and the aggregated NR-V2X data rate R_v are traced in ns-3 and are provided as input to the DRL agent as its state space. The algorithm is implemented, and the DRL agent is trained according to equations (14) and (15). The policy and target networks in the DRL system consist of an input layer, two fully connected layers, and an output layer. Each network includes two fully connected layers comprising 512 and 256 neurons. The Adamax optimizer is employed, and parameter optimization is performed using the SGD algorithm. The number of iterations or steps within epochs may be reduced if early stopping is triggered or if there is no significant change in the reward for more than five iterations; a changing threshold β is established for this purpose. The simulation parameters for the DRL approach are crucial for optimizing the learning process and overall performance. The discount factor (γ) is set to 0.04, which influences the weight given to future rewards in the learning process, emphasizing short-term gains. The learning rate (α) is 0.5, determining the step size at each iteration while moving toward a minimum of the loss function, hence balancing the trade-off between speed and accuracy of convergence. The learning process spans around 100 epochs until the probability $\mathbf{E} \in (0, 1)$ reaches a higher value (0.8) when the agent exploits the learned correlations in the state space. Additionally, the epsilon decay (ϵ) is initially configured at 0.6, which controls the rate at which the exploration probability decreases, balancing exploration and exploitation over time. Lastly, the number of iterations (K) is set to 10, specifying the number of times the training

TABLE 2: Comparison of the proposed approach with the other existing approaches.

Algorithm/Technique	Coexistence System with Wi-Fi	Aggregated Wi-Fi Throughput (Mbps)	Aggregated V2X Throughput (Mbps)
Traditional [16]	LTE-V2X mode - 3	8	22
Average [16]	LTE-V2X mode - 3	9	28
LBT [6]	LTE-V2X mode - 3	13	13
Q-Learning [16]	LTE-V2X mode - 3	14	28
Static ($T_v = 2.5$)	NR-V2X mode - 1	29.14	33.72
Static ($T_w = 2.5$)	NR-V2X mode - 2	29.14	31.392
DRL (proposed)	NR-V2X mode-1	**36.03**	**41.63**
DRL (proposed)	NR-V2X mode-2	**36.13**	**38.03**

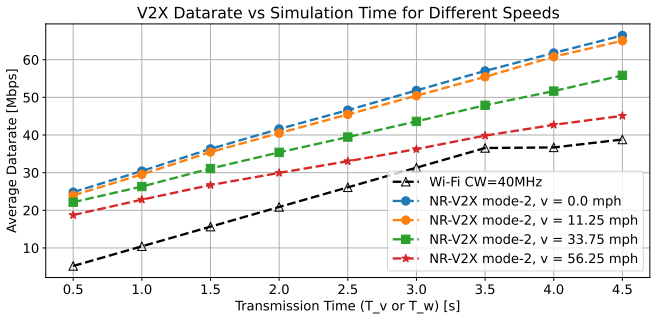


FIGURE 6: Analysing the datarate with respect to the given transmission time for NR-V2X mode -2 and Wi-Fi 6, both stand-alone systems with the impact of mobility on the NR-V2X system.

loop is executed per epoch. These parameters collectively ensure the DRL model is well-tuned for effective learning and performance.

A standard Epsilon greedy policy is applied to balance exploration and exploitation by the agent. The value of ϵ is initialized at 0.6 and subsequently reduced as follows:

$$\epsilon_{t+1} = \frac{\epsilon_t - \epsilon_{min}}{K} \tag{18}$$

where the ϵ_t is the value of ϵ in the iteration 't' and ϵ_{min} is the minimum value of epsilon recorded in that particular epoch (Algorithm 1 line 26).

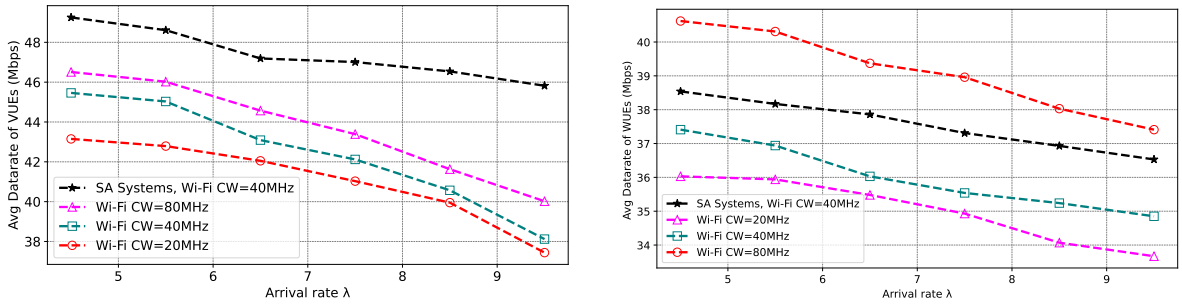
We consider two performance metrics. First is throughput, which evaluates and compares the model performance with the existing approaches. The second is fairness, which measures the impact of varying channel conditions on both subsystems. We use the formulation for fairness in (19), which is the modified version of Jain's fairness index. The typical Jain's fairness index calculates the fairness at the node level. However, for this problem, we need to consider fairness at the system level.

$$F = \frac{\left(\sum_{t=1}^{T_{\max}} \left(\sum_{i=1}^{WUEs} R_{w,i,t} + \sum_{i=1}^{VUEs} R_{v,i,t}\right)\right)^2}{(N_{total})\left(\sum_{t=1}^{T_{\max}} \left(\sum_{i=1}^{WUEs} R_{w,i,t}^2 + \sum_{i=1}^{VUEs} R_{v,i,t}^2\right)\right)} \tag{19}$$

The above equation calculates the system level fairness, where $R_{v,i,t}$ and $R_{w,i,t}$ are the throughputs of the V2X and Wi-Fi users at iteration t, and $N_{total} = VUEs + WUEs$. This system-level fairness index aims to measure if fair transmission time is allocated to Wi-Fi and V2X users by observing their data rates. Additionally, we note that for this specific problem, learning short-term dynamics is more important. Increasing the discount factor makes the gradient learn long-term dynamics, yielding poor results. Thus, learning a shorter horizon for capturing recent traffic trends is more effective.

D. Simulation Results

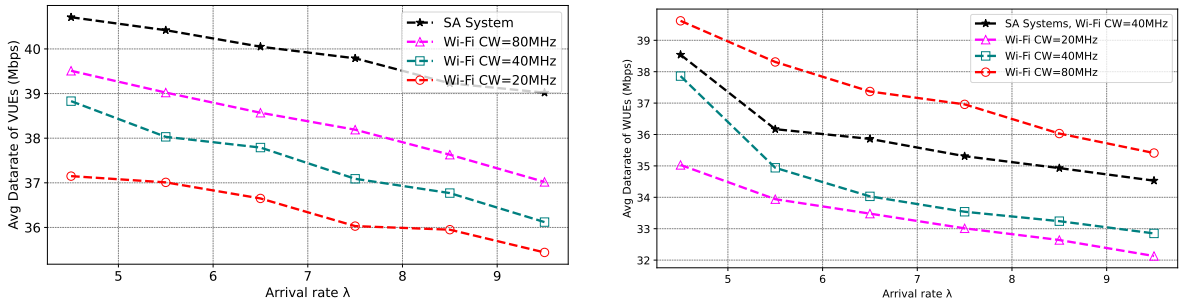
In this section, we present a comparative analysis of the results from the proposed approach with various existing schemes: Traditional and average methods referred in [16], LBT [6], and Q-Learning [16] when applied to the LTE-V2X and Wi-Fi coexistence system. Table - 2 compares the 5G NR-V2X and Wi-Fi coexistence system with the LTE-V2X and Wi-Fi coexistence scenarios. The effectiveness of 5G technology when integrated with the proposed approach, particularly in the case of limiting transmission time, is evident from the significantly enhanced aggregated data rates observed for VUEs, i.e., 41.63 Mbps and 38.03 Mbps for modes 1 and 2, respectively. It outperforms the static method (same system, but the action is static $T_w/T_v = 2.5$) applied to NR-V2X and Wi-Fi coexistence, and it also performs better than the previously proposed works on LTE-V2X and Wi-Fi coexistence [16]. First, we analyze the performance of both Stand-alone Wi-Fi and NR-V2X by carrying out the simulations for a $T_v = T_f = [0, ...5]$. This will help us understand the convergence limit of the algorithms, as the simulations will let us observe the maximum possible throughput for both sub-systems. Figure 6 shows the achievable throughput for different transmission periods. The Figure 6 also shows the impact of relative velocity 'v' between the VUEs. Thus, with the increase in relative velocity, the Average data rate reduces. TO maintain a more realistic scenario, we effectively maintain the relative velocity of 56.25 mph while simulating the NR-V2X scenario in the coexistence system. Additionally, the curve is increasing with transmission periods. We consider the $VUEs = WUEs = 6$ over a channel of 40MHz. Wi-Fi is moving over 31 Mbps when $T_w = 3$, and for NR-V2X, it is moving over 32 Mbps when $T_v = 3$. These are some critical points to be observed, as when Wi-Fi gets a transmission time greater than 2.5, it gets priority over the NR-V2X system, and vice versa. Suppose the coexisting system's average data rate (Wi-Fi or NR-V2X) comes close to these critical values. In that case, we can assume that our algorithm is performing well while



(a) Average data rate of VUEs vs. Arrival rate (λ) of VUEs and WUEs.

(b) Average data rate of WUEs vs. λ of VUEs and WUEs.

FIGURE 7: Analysing the effect of using different CW's on data rates of VUEs (V2X mode 1 SL) and WUEs. Note that 'SA' stands for stand-alone Wi-Fi/NR-V2X Systems.



(a) Average Data rate of VUEs vs. Arrival rate (λ) of VUEs and WUEs

(b) Average Data rates of WUEs vs. Arrival rate (λ) of VUEs and WUEs

FIGURE 8: Analyzing the effect of using different CWs on data rates of VUEs (V2X mode - 2) and WUEs. Note that 'SA' stands for stand-alone Wi-Fi/NR-V2X Mode 2 systems.

reaching convergence, as the Wi-Fi/V2X system performs equivalent to its stand-alone system with an average of 6 users, and a prioritized transmission time. We will use these later to analyze the performance of the proposed coexistence approach.

With the proposed method, the Wi-Fi throughput of 36.03 and 36.13 in the case of coexistence with NR-V2X modes 1 and 2, respectively, outperforms the static methods and the previously proposed methods for LTE-V2X and Wi-Fi coexistence. Comparing the results, we observe a substantial improvement of around 24% and 23% in the average data rates for VUEs and Wi-Fi users (WUEs), respectively, when employing DRL on NR-V2X and Wi-Fi coexistence, as compared to the static methods. There is an improvement of 48% in NR-V2X throughput when compared to the existing RL-based LTE-V2X and Wi-Fi coexistence and other static approaches. The DRL approach is capable of controlling transmission time after learning about Wi-Fi and V2X traffic. This improvement substantiates the assertion that the 5G NR-V2X technology, which can achieve better throughput within reduced transmission times, seems more reliable than LTE-V2X in addressing high data transmission demands in ITS. To the best of our knowledge, no existing work focuses on NR-V2X (both modes) and Wi-Fi 6E coexistence leveraging RL-based or any other traditional methods. Due to the lack of LTE-V2X simulations with a similar scenario to that of 5G NR-V2X in ns-3, we couldn't apply and compare the other state-of-the-art methods proposed for the LTE-V2X and Wi-Fi coexistence. Now, as discussed before, we can compare the results with those in Figure 6. We can conclude that both systems can achieve the average data rate in the coexistence scenario (Figure 8), which is nearer to that in the stand-alone scenario with transmission time of 3 seconds (Figure 6). Since the total frametime in the coexistence scenario is 5 seconds, it's a win-win case for both the systems if they can get an average datarate equivalent to or above that at the transmission time of 3 seconds in the stand-alone scenario. Thus, we can justify the convergence of the model as it can reach the performance equivalent to the stand-alone systems scenario. With the improved results, we can notice that the proposed learning mechanism effectively makes the agent learn the best policy based on the varying VUEs and WUEs traffic. It can learn the optimal parameters of the policy network and hence learn how to optimally address the traffic on both subsystems and manage the transmission time of the NR-V2X subsystem accordingly. We further validate the coexistence simulations by varying system parameters. First, we vary the VUEs and WUEs arrival rates (equal to λ). This can be observed in Figs. 7-a and 7-b, the V2X and Wi-Fi throughput decline as the arrival rate increases. The results

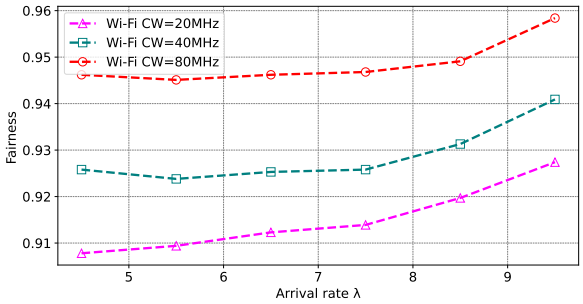


FIGURE 9: Analyzing the effect of different CWs for Wi-Fi on system fairness while coexisting with NR-V2X mode-1.

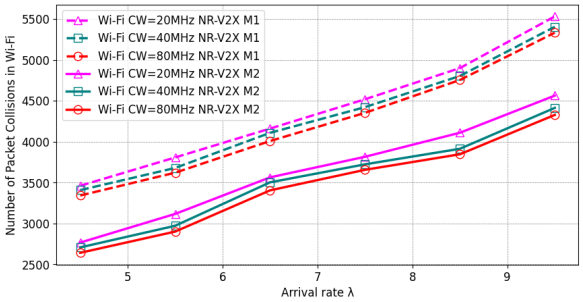


FIGURE 10: Number of Wi-Fi packet collisions vs. λ by varying the CWs while coexisting with NR-V2X mode - 1 (M1) and mode - 2 (M2).

are generated by simulating with a trained DRL agent for 20 epochs and averaging the throughput.

We vary the CW of the Wi-Fi system and analyze the effect on the NR-V2X and Wi-Fi performance. With a higher CW, the Wi-Fi channel will cover a larger span, and hence, the overall throughput will increase. In Figure 7-b, we can notice that the Wi-Fi throughput curve shifts above as we increase the CW from 20 MHz to 40 and 80 MHz. The 80 MHz scenario even outperforms the stand-alone[2] Wi-Fi as the frequency resources are extensive. Simultaneously, in Figure 7-a, the V2X throughput is also increased. This shows that the DRL agent allocates less time to Wi-Fi if it has higher bandwidth. It can be observed in Figs. 8-a and 8-b, as the arrival rate increases, the data rates for both V2X and Wi-Fi decline. We also vary the CW of the Wi-Fi system and analyze its impact on NR-V2X and Wi-Fi performance. As the CW increases, the Wi-Fi channel spans a more extensive frequency range, increasing overall throughput. In Figure 8-b, it is evident that the Wi-Fi throughput improves as the CW is increased from 20 MHz to 40 MHz and 80 MHz. However, in Figure 8-a, the V2X throughput increases with increasing Wi-Fi CW, indicating that the DRL agent allocates less time to Wi-Fi when higher bandwidth is available to the Wi-Fi system.

Figure 9 shows that fairness is improved with many users as the arrival rate λ of VUEs and WUEs increases. This indicates that the V2X throughput has declined with a higher slope than the Wi-Fi throughput and came closer to the Wi-Fi throughput. Additionally, fairness is also improved in the case of higher Wi-Fi CWs. Figure 10 illustrates the effect of varying arrival rates λ and CW on the number of packet collisions in Wi-Fi when they coexist with two NR-V2X modes. It is observed that as the arrival rate increases, the number of packet collisions in Wi-Fi rises for all configurations. Wi-Fi systems with a CW of 20 MHz experience the highest collisions, whereas those with 80 MHz CW show the lowest collisions across both modes. Notably, when coexisting with NR-V2X mode-2, Wi-Fi consistently results in higher packet collisions than Mode 1 for the same CW and arrival rate. This indicates that NR-V2X mode-1 and Wi-Fi systems are more collision-prone environments. This also suggests that mode-1 requires higher spectrum coverage than mode-2 as it requires higher transmission time, degrading Wi-Fi performance. The fact is reasonable as it has a more advanced channel access mechanism than mode-2, i.e., the DG scheduling under the network coverage. Moreover, the results indicate that mode-2 communication requires less transmission time than mode-1, making the system fairer for Wi-Fi users.

The DRL model's policy network and target networks consist of an input layer that takes the current NR-V2X and Wi-Fi system state, followed by two fully connected Dense layers that extract and refine features using ReLU activation. The final output layer computes Q-values for all possible actions, enabling the agent to select the optimal transmission time. The policy network's computational complexity is calculated in terms of Floating Point Operations (FLOPs) as 36,608 FLOPs. When benchmarked on a machine with 8 GB RAM and an Intel Core i7 CPU at 2.2 GHz, it achieves an average inference time of 9.364 ms. In a real-time scenario, the algorithm will be deployed on the gNB in case of Wi-Fi coexistence with NR-V2X mode-1, and on the Wi-Fi AP in case of coexistence with NR-V2X mode-2. At any infrastructure, just involving an 8 GB RAM microprocessor of 2.2 GHz, internally inside the infrastructure of gNB/AP or an external system (like Raspberry Pi 5) will provide the inference time of around 9.364 ms. Given that the typical time required for a successful transmission in the NR-V2X and Wi-Fi coexistence scenario is significantly higher, the proposed model can set the system parameters in less than $\frac{1}{5}^{th}$ of this duration. This rapid inference ensures seamless coexistence between NR-V2X and Wi-Fi systems without introducing additional latency.

V. Discussions

Through the extensive simulations in the ns-3, we created an example of designing a frame-based Media Access Control (MAC) protocol that can be created and managed algorithmically to make the two subsystems, NR-V2X and Wi-Fi, optimally coexist. However, we observed a few shortcomings in the approach while performing the simulations. There are a few iterations where we observed that the data rate of the NR-V2X subsystem drops below 20 Mbps when the number of VUEs is still high. This indicates that an individual VUE will receive lower throughput, and some VUEs may be deprived of sufficient resources. These are unfavorable events

[2]By stand-alone mode, we mean the scenario of no coexistence (only Wi-Fi/NR-V2X)

for the 5G NR-V2X systems because, as per the 3GPP Rel 15, the C-V2X is developed for supporting the ITS services, and they might be supporting/implementing critical services such as road safety. A sudden drop in NR-V2X throughput can lead to critical failures. To prevent this, transmission time allocation must ensure that no VUE experiences throughput below a set threshold. This necessitates a constrained optimization framework that enables Deep Q-Learning to learn an optimal policy while adhering to this safety constraint.

While the proposed framework has been extensively evaluated using ns-3 simulations, its scalability and applicability in larger and more diverse network environments remain important considerations. Interference management may be crucial to address, as hidden node problems cannot be addressed in this study. Transmissions from a hidden Wi-Fi node can lead to inter-system interference when a V2X transmission is ongoing. Furthermore, deploying a Deep-RL model in real-time wireless environments can pose additional challenges. The time and computational complexity should be under a certain limit to ensure timely responses, so there would not be latency concerns. Therefore, it is essential to create lightweight Reinforcement Learning models for prompt decision-making. Additionally, the Deep-RL agent in the exploration state can make incorrect decisions that can negatively impact the throughput of any coexisting systems. For example, while exploring, the agent allocates very little transmission time to the NR-V2X system in a high-traffic scenario. As we know, the NR-V2X is a standard proposed for implementing safety-critical ITS use cases, and having a network outage can have a critical impact on the running applications. Thus, a constraint can be set to avoid such situations. Implementing Constrained RL or Safe-RL could resolve such issues. Future work should evaluate the framework in more complex network settings, including large-scale vehicular testbeds, to assess its robustness under real-world constraints. Addressing these scalability challenges looks crucial for transitioning from simulation-based validation to practical deployments.

VI. Conclusion

This work proposes a novel frame-based MAC protocol for NR-V2X and Wi-Fi coexistence. Unlike static methods, we model it as a decision-making problem and apply DRL to dynamically control transmission times based on real-time network conditions. We model the DRL scenario using the OpenAIGym framework. Choosing the gNB as the agent in the case of coexistence with mode-1 and the Wi-Fi AP in the case of mode-2, the approach is executed using the ns-3 simulator by developing the two above-mentioned subsystems. The simulation results conclude that the DRL-based agent maintains a reasonable transmission time control, validated by the results in Table 2, with high-performance improvements of the proposed approach compared with the existing schemes. The findings indicate that utilizing DRL for NR-V2X and Wi-Fi coexistence results in significant improvements in the average data rates for VUEs and WUEs with an increase of approximately 24% and 23%, respectively when compared to the static method for coexistence. Our analysis shows that Wi-Fi performance degrades more when coexisting with NR-V2X mode-1 than mode-2, as mode-1, operating under network coverage, requires more spectrum resources.

VII. Acknowledgement

This research was conducted at the Machine Intelligence Computing and xG Networks (MICxN) Lab, with infrastructural support from the School of Engineering and Applied Science, Ahmedabad University, and Boston College. The work is supported by the Department of Science and Technology, Gujarat Council of Science and Technology (DST-GUJCOST): GUJCOST/STI/R&D/2024-25/3893.

REFERENCES

[1] K. Sehla, T. M. T. Nguyen, G. Pujolle, and P. B. Velloso, "Resource allocation modes in C-V2X: From LTE-V2X to 5G-V2X," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8291–8314, 2022.

[2] FCC, "First report and order, further notice of proposed rulemaking, and order of proposed modification, https://docs.fcc.gov/public/attachments/fcc-20-164a1.pdf," Nov. 2020.

[3] European Commission, "Mandate to CEPT to study feasibility and identify harmonized technical conditions for wireless access systems including radio local area networks in the 5925-6425 MHz band for the provision of wireless broadband services," Dec. 2017.

[4] IEEE, "IEEE P802.11ax/D6.1; Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Amendment 1: Enhancements for High Efficiency WLAN," May 2020.

[5] R. Zhang, M. Wang, L. X. Cai, X. Shen, L.-L. Xie, and Y. Cheng, "Modeling and analysis of MAC protocol for LTE-U co-existing with Wi-Fi," in *Proc. of IEEE GLOBECOM*, 2015, pp. 1–6.

[6] H. Ko, J. Lee, and S. Pack, "A fair listen-before-talk algorithm for coexistence of LTE-U and WLAN," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10 116–10 120, 2016.

[7] V. Sathya, M. Mehrnoush, M. Ghosh, and S. Roy, "Analysis of CSAT performance in Wi-Fi and LTE-U coexistence," in *Proc. of IEEE ICC Workshops*, 2018, pp. 1–6.

[8] 3GPP, "3rd generation partnership project; technical specification group services and system aspects; release 16 description; summary of rel-16 work items," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 21.916 V16.2.0, June 2022.

[9] G. Naik and J.-M. J. Park, "Coexistence of Wi-Fi 6E and 5G NR-U: Can we do better in the 6 GHz bands?" in *Proc. of IEEE INFOCOM*, 2021, pp. 1–10.

[10] Y. Kakkad, D. K. Patel, S. Kavaiya, S. Sun, and M. López-Benítez, "Optimal 3GPP fairness parameters in 5G NR unlicensed (NR-U) and WiFi coexistence," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 5373–5377, 2023.

[11] R. Yarinezhad, E. Ekici, M. I. Khan, T. Shimizu, and O. Altintas, "A decentralized coexistence protocol for V2X and Wi-Fi in Unlicensed bands," in *Proc. of IEEE VNC*, 2024, pp. 65–72.

[12] Z. Wu, S. Bartoletti, V. Martinez, V. Todisco, and A. Bazzi, "Analysis of co-channel coexistence mitigation methods applied to IEEE 802.11p and 5G NR-V2X sidelink," *Sensors*, vol. 23, no. 9, 2023.

[13] M. Noor-A-Rahim, Z. Liu, H. Lee, M. O. Khyam, J. He, D. Pesch, K. Moessner, W. Saad, and H. V. Poor, "6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities," *Proceedings of the IEEE*, vol. 110, no. 6, pp. 712–734, 2022.

[14] N. Rupasinghe and Güvenç, "Reinforcement learning for licensed-assisted access of LTE in the unlicensed spectrum," in *Proc. of IEEE WCNC*, 2015, pp. 1279–1284.

[15] Y. Su, X. Du, L. Huang, Z. Gao, and M. Guizani, "LTE-U and Wi-Fi Coexistence Algorithm Based on Q-Learning in Multi-Channel," *IEEE Access*, vol. 6, pp. 13 644–13 652, 2018.

[16] Y. Su, M. LiWang, Z. Gao, L. Huang, S. Liu, and X. Du, "Coexistence of Cellular V2X and Wi-Fi over Unlicensed Spectrum with Reinforcement Learning," in *Proc. of IEEE ICC*, 2020, pp. 1–6.

[17] J. Tan, L. Zhang, Y.-C. Liang, and D. Niyato, "Deep reinforcement learning for the coexistence of LAA-LTE and WiFi systems," in *Proc. of IEEE ICC*, 2019, pp. 1–6.

[18] E. Pei, Y. Huang, L. Zhang, Y. Li, and J. Zhang, "Intelligent access to Unlicensed Spectrum: A mean field based deep reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2325–2337, 2023.

[19] B. A. G. Sutton, Richard S, *Reinforcement learning: An introduction*. MIT press, 2018.

[20] A. Plaat, *Deep Reinforcement Learning*. Springer Nature Singapore, 2022.
[21] D. Magrin, S. Avallone, S. Roy, and M. Zorzi, "Validation of the ns-3 802.11ax OFDMA implementation," in *Proceedings of the 2021 Workshop on ns-3*, 2021, pp. 1–8.
[22] CTTC, "NR-V2X tutorial," 2022, presented at the WNS3-2022. [Online]. Available: https://www.nsnam.org/tutorials/consortium22/WNS3-2022-NR-V2X-Tutorial-Zoraze-Ali.pdf

Kashish D. Shahreceived a BTech. in Information and Communication Technology from the School of Engineering and Applied Sciences, Ahmedabad University, Gujarat, India, in 2022. He is currently a full-time Ph.D. Scholar in Computer Science and Engineering Department at Ahmedabad University. He served as a Junior Research Fellow in the 5G-ITS project financed by DST (Department of Science and Technology, India)-GUJCOST. CUrrently, he is a JRF under a DoT 6G project focusing on developing 5G NR-V2X testbed. His current research is focused on Integrated Sensing and Communication (ISAC) and AI-based Coexistence of 5G NR-V2X and WiFi systems under the unlicensed spectrum. His research interests include Bio-Medical Image Segmentation, Computer Vision, Reinforcement Learning, Deep Learning, Wireless Networks, and Cellular-V2X.

Dhaval K. Patel(Senior Member, IEEE) received the B.E. and M.E. degrees (Hons.) in communication systems engineering from Gujarat University, Ahmedabad, India, in 2003 and 2010, respectively, and the Ph.D. degree in electronics and communications from the Institute of Technology, Nirma University, Ahmedabad, in 2014. He was a Visiting Faculty at the Franklin W. Olin College of Engineering-Massachusetts, Needham, MA, USA. From 2011 to 2014, he was a Junior Research Fellow at the Post Graduate Laboratory for Communication Systems, Nirma University. Since 2014, he has been with the School of Engineering and Applied Science, Ahmedabad University, India, where he is currently an Associate Professor. His research interests include vehicular cyber-physical systems, 5G wireless networks, non-parametric statistics, and physical layer security. He is the Principal Investigator of Research Projects funded by Telecom Center of Excellence (TCoE) - Department of Telecommunications (DoT), the Department of Science and Technology, U.K. India Education and Research Initiative (UKIERI), Association of Southeast Asian Nations (ASEAN) India Collaborative Research and Development Project, and the Gujarat Council on Science and Technology.

Brijesh Sonireceived the B.E. in Electronics Engineering and the M.E. in Communication Systems Engineering from Gujarat Technological University in 2014 and 2017, respectively, and the Ph.D. degree from the School of Engineering and Applied Science, Ahmedabad University in July 2021. During 2017-2020, he was also associated as a Junior Research Fellow at Ahmedabad University for the Department of Science and Technology (DST) — U.K.India Education and Research Initiative (UKIERI) research project jointly funded by DST and British Council, U.K. He was a postdoctoral researcher at Boston College, MA, USA from 2022-2024. Since August 2024, he is a Professional Practice Assistant Professor in the Department of Computer Science and Engineering at The Ohio State University, OH, USA. His research interests include cognitive radio networks, and applied machine learning/deep learning for xG wireless networks.

Siddhartan Govindasamy(Member, IEEE) received the bachelor's, master's, and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), in 1999, 2000, and 2008, respectively. At MIT, he worked on a master's thesis on speech enhancement algorithms in partnership with Qualcomm Inc. and a Ph.D. thesis on ad-hoc wireless networks with multi-antenna devices. From 2000 to 2003, he was a DSP Engineer and later a Senior DSP Engineer at Aware Inc. where he worked on developing broadband modem technology. From 2008 to 2020, he was an Assistant Professor and later an Associate Professor of electrical and computer engineering at the Olin College of Engineering, Needham, MA, USA, where he conducted research on large multiple-input multiple-output (MIMO) systems and optical wireless communications. He joined the Department of Engineering, Boston College, as a founding faculty member, in fall 2020, where he is currently a Sabet Family Dean's Faculty Fellow and Professor of Engineering. He is a coauthor of the textbook Adaptive Wireless Communications: MIMO Channels and Networks (Cambridge University Press, 2013). His research interests include in large MIMO systems and signal detection and processing.

Mehul Raval Ph.D. (University of Pune), is a Professor at Ahmedabad University with over 27 years of academic experience. He has held key positions at SCET Surat, DA-IICT Gandhinagar, PDEU Gandhinagar (as founding head of ICT), and Ahmedabad University. Dr. Raval's research spans computer vision, image processing, machine learning, and engineering education, resulting in numerous publications and funded projects. He serves on editorial boards of leading AI journals and has held leadership roles in IEEE Gujarat. His international engagements include research and visiting professorships in Japan and the USA, mentoring students, and advancing curriculum development.

Mukesh Zaveriis serving as professor in the Department of Computer science and engineering at Sardar Vallabhbhai National Institute of Technology Surat, Gujarat, India since 1993. His research areas are image processing, wireless network, and machine learning. He obtained his doctoral degree from the Indian Institute of Technology - Bombay.