

# Spectrum Sharing between Cellular and Wi-Fi Networks Based on Deep Reinforcement Learning

Bayarmaa Ragchaa and Kazuhiko Kinoshita

Department of Information Science and Intelligent Systems,  
Tokushima University, Japan

## **ABSTRACT**

Recently, mobile traffic is growing rapidly and spectrum resources are becoming scarce in wireless networks. Due to this, the wireless network capacity will not meet the traffic demand. To address this problem, using cellular systems in an unlicensed spectrum emerged as an effective solution. In this case, cellular systems need to coexist with Wi-Fi and other systems. For that, we propose an efficient channel assignment method for Wi-Fi AP and cellular NB, based on the DRL method. To train the DDQN model, we implement an emulator as an environment for spectrum sharing in densely deployed NB and APs in wireless heterogeneous networks. Our proposed DDQN algorithm improves the average throughput from 25.5% to 48.7% in different user arrival rates compared to the conventional method. We evaluated the generalization performance of the trained agent, to confirm channel allocation efficiency in terms of average throughput under the different user arrival rates.

## **KEYWORDS**

Wi-Fi, spectrum sharing, unlicensed band, deep reinforcement learning

## **1. INTRODUCTION**

The amount of mobile data traffic is growing at an annual rate of around 54% in 2020-2030. Furthermore, the global mobile traffic per month would then be estimated to reach 543EB in 2025 and 4394EB in 2030 [1]. Under these predictions, the wireless network capacity will not meet the exponential growth of the mobile traffic demand.

To tackle this problem, extending LTE and 5G to unlicensed spectrum has emerged as a promising and effective solution that can assist in exploiting the wireless spectrum **more efficiently** and can also be a good neighbor with the other occupants [2], [3]. In principle, 5G New Radio Unlicensed (NR-U) systems are allowed to operate in any unlicensed bands (from 1 to 100GHz) [4], but the initial industry focus is on the 5 GHz bands. Also, [5] expects both License Assisted Access (LAA) and NR-U to coexist in 5GHz unlicensed bands in future years. There **are** up to 500 MHz of spectrum in this band that is available on a global basis for unlicensed applications [6], [7].

When different technologies operate on the same band without any coordination, however, it causes a significant interference that reduces the average throughput per user. [8] showed that in the absence of any cooperation technique in the LAA/Wi-Fi heterogeneous networks for the same frequency band, the user throughput of Wi-Fi had a 96.63% of decrease, whereas user throughput of LTE was slightly affected by 0.49%, compared to the case in which both technologies operating alone. In this regard, several significant works have **been** proposed for coexistence between LTE-U and Wi-Fi by Carrier Sensing Adaptive Transmission (CSAT)

[3],[4], Listen Before Talk (LBT) [9], or Almost Blank Subframe (ABS) [8]. In common, they allow LTE and Wi-Fi systems to share the unlicensed band by checking the availability of the channel before transmission. Therefore, there is sufficient work to investigate the coexistence of LTE and Wi-Fi technologies in unlicensed spectrum bands based on traditional methods.

[10] indicates many limitations for traditional optimization approaches in wireless resource allocation problems. In other words, traditional methods are used to solve Radio Resource Allocation and Management (RRAM) optimization problems that require complete or quasi-complete knowledge (difficult/impossible to obtain this information in real-time) of the wireless environment, such as accurate channel models and real-time channel state information. Moreover, traditional methods are mostly computationally expensive and cause notable timing overhead. This shows them inefficient for most emerging time-sensitive applications. To overcome those limitations, Machine Learning (ML) based methods, especially Deep Reinforcement Learning (DRL) can be an effective solution and take judicious control decisions with only limited information about the network statistics [11]. There are three ways, including supervised learning, objective-oriented unsupervised learning, and reinforcement learning paradigms to incorporate DL in solving optimization problems. From these methods, we have selected to use the DRL approach for the efficient channel assignment problem. Note that there is no explicit output in our problem as a ground truth label for the training model. In this case, we consider two unknown metrics which are channel assignment pattern and average throughput for each AP/NBs in our assumed environment. For that, the reinforcement learning method can be applied as an effective solution for these two unknown metrics, where action and reward can represent channel allocation information and average throughput, respectively. The obtained results indicate that our proposed method provides major improvements in average throughput in the developed environment compared to traditional methods.

In this work, we propose to maximize the average throughput by assigning suitable channels to Wi-Fi Access Points (APs) and cellular BSs. Specifically, we developed an emulator for training a DQN agent which includes densely deployed APs and eNodeBs (eNBs) in a rectangular area. First, in order to apply the DRL method, the state information of the assumed environment is converted to the MDP framework. When training the DQN agent only one channel state of AP or eNB is changed during each episode. For these generated states an action is tried step by step according to the epsilon greedy algorithm. Hence, random actions in the first phase of training DQN and the final phase of the training process greedy actions are offered for the observed states. Accordingly, the trained agent is able to assign the optimal channel to each AP/eNB based on the learned knowledge of the environment which includes information on the user's variation and channel state. On the other hand, if they receive the highest reward based on learned knowledge for each time step in an episode, the agent can select optimal action for each state according to the rule of the epsilon greedy algorithm. In our case, it means that the optimal channel is assigned to each AP/eNB based on the highest average throughput. Consequently, the expected metrics such as average throughput is possible to enhance for each AP and eNB in our assumed heterogeneous network. It can assist in the more efficient management of the wireless spectrum for the ever-increasing wireless traffic.

The rest of this paper is organized as follows. We survey some related works in [Section 2](#); [Section 3](#) presents the proposed method and the general system architecture of the assumed environment. [Section 4](#) presents the performance evaluation of the proposed method. Finally, [Section 5](#) concludes this paper and future work.

## 2. RELATED WORKS

A cellular system operating in unlicensed spectrum bands has emerged as a promising and effective solution to meet the ever-increasing traffic growth that can assist in exploiting the wireless spectrum in a more efficient way [3].

## 2.1 Spectrum sharing techniques

Extending cellular systems such as LTE and 5G into unlicensed bands, currently dominated by Wi-Fi (IEEE 802.11-based technologies), brings about challenges related to regulatory requirements including spectrum sharing, a maximum channel occupancy time, a minimum occupied channel bandwidth, and fair coexistence with incumbent systems [12]. Therefore, it is not trivial for cellular and Wi-Fi to coexist as-is due to the differences in their Medium Access Control (MAC) protocols. One MAC protocol cannot satisfy all the requirements of various kinds of applications because the different kinds of protocols assume different hardware, and applications [34]. A cellular system uses a centralized channel access mechanism based on Orthogonal Frequency Division Multiple Access (OFDMA). In contrast, Wi-Fi uses a distributed channel access mechanism based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). In particular, LTE transmits according to predefined schedules, whereas Wi-Fi is governed by a CSMA protocol, by which stations transmit only when sensing the channel idle. Due to these fundamental differences between the two access systems, of which LTE is more aggressive, i.e., LTE unlicensed will create harmful interference to Wi-Fi [9].

To address this issue, a coexistence mechanism is required to manage the interference between two different technologies. To this end, several coexistence mechanisms including LBT, CSAT and ABS have been developed into the same channel-sharing methods in unlicensed bands, for legacy (LAA and LTE-U) of eNR. Above all, LBT is the most popular coexistence mechanism [14].

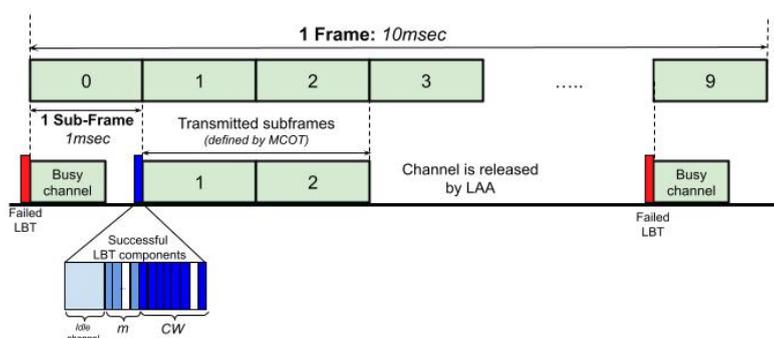


Figure 1. The basic LBT-based channel access mechanism

The main principle of LBT (as represented by the blue color, in figure 1) is defined as follows:

- A transmitter before starting a transmission, first waits for the channel to be idle for  $16\mu\text{s}$ .
- The device performs Clear Channel Assessment (CCA) after each of the ' $m$ ' required observation slots.
- For the back-off-stage, the device selects a random integer  $N$  in  $\{0, \dots, CW\}$ , where  $CW$  is the contention window. CCA is performed for each observation slot and results either in decrementing  $N$  by 1 or freezing the backoff procedure. Once  $N$  reaches 0, a transmission may commence.
- The length of the transmission is upper bounded by the Maximum Channel Occupancy Time (MCOT) up to 10ms.
- If the transmission is successful, the responding device may send an immediate acknowledgment (without a CCA) and reset  $CW$  to  $CW_{min}$ . If the transmission fails, the  $CW$  value is doubled (up to  $CW_{max}$ ) before the next retransmission [15].

Furthermore, when different technologies share the same band in heterogeneous networks, especially in densely deployment scenarios, there is a significant interference that reduces the system performance including user's throughput. To address this problem, a central controller is introduced to manage both APs and eNBs in a centralized manner to improve the system performance.

## 2.2 Spectrum sharing between Wi-Fi and cellular without ML

Several survey and tutorial papers [6],[8],[13],[16] analyzed overall issues which are related to spectrum-sharing and the coexistence of Wi-Fi and LTE-U/NR-U technologies from different aspects. For example, [9] systematically explores the design of efficient spectrum-sharing mechanisms for inter-technology coexistence in a system-level approach, by considering the technical and non-technical aspects in different layers. Furthermore, [4] provides a comprehensive survey on full spectrum sharing in cognitive radio networks including the new spectrum utilization, spectrum sensing, spectrum allocation, spectrum access, and spectrum hand-off towards 5G. [6] addresses coexistence issues between several important wireless technologies such as LTE/Wi-Fi as well as radar operating in the 5GHz bands. [8] investigates coexistence-related features of Wi-Fi and LTE-LAA technologies, such as LTE carrier aggregation with the unlicensed band, LTE and Wi-Fi MAC protocols comparison, coexistence challenges and enablers, the performance difference between LTE-LAA and Wi-Fi, as well as co-channel interference. [14] investigates genetic algorithm-based channel assignment and access system selection methods in densely deployed LTE/Wi-Fi integrated networks to improve the user throughput and fairness issue. [15] evaluates the impact of LAA under its various QoS settings on Wi-Fi performance in an experimental testbed. Various methods were proposed in [17] to adapt the transmission and waiting times for LAA based on the activity statistics of the existing Wi-Fi network which is exploited to tune the boundaries of the CW. This method provides better total aggregated throughput for both coexisting networks compared to the 3GPP cat4 LBT algorithm. [12] investigates the user level performance attainable over LTE/Wi-Fi technologies when employing different settings, including LTE duty cycling patterns, Wi-Fi offered loads, transmit power levels, modulation, and coding schemes, and packet sizes. The interference impact of LAA-LTE on Wi-Fi is studied in [5] under various network conditions using experimental analysis in an indoor environment. [7] presents a coexistence study of LTE-U and Wi-Fi in the 5.8GHz unlicensed spectrum based on the experimental testing platform which is deployed to model the realistic environment.

Analytical models are established in several studies [18-20] to evaluate the downlink performance of coexisting LAA and Wi-Fi networks by using the Markov chain. Particularly, [21] establishes a theoretical framework based on Markov chain models to calculate the downlink throughput performance of LAA and Wi-Fi systems in different coexistence scenarios. In recent years, the coexistence between Wi-Fi and LTE systems has been sufficiently studied for the 5GHz unlicensed band. NR-U is a successor to 3GPP's Release 13/14 LTE-LAA [16]. Therefore, initially, NR-U is expected to coexist with Wi-Fi and LTE-LAA technologies in the 5GHz unlicensed spectrum band. [22] proposes a fully blank subframe-based coexistence mechanism and derives optimal air time allocations to cellular/IEEE 802.11 nodes in terms of blank subframes for 5G NR-U operating in both the licensed and unlicensed mmW spectra for in-building small cells. Furthermore, [23] presents a system-level evaluation of NR-U and Wi-Fi coexistence in the 60GHz unlicensed spectrum bands based on a competition-based deployment scenario.

## 2.3 Spectrum sharing between Wi-Fi and cellular with ML

So far, a large number of studies are addressed the coexistence between cellular and Wi-Fi technologies without ML. During the last few years, ML and DL-based methods [11], [24], [25], [26] are proposed for the communication systems problem, such as spectrum sensing, spectrum allocation, spectrum access, etc. Particularly, [24] proposes a DRL-based channel allocation scheme that enables the efficient use of experience in densely deployed wireless local area networks. The existing works for the CSAT/CA mechanism in LTE-U/Wi-Fi heterogeneous networks mostly focus on the power control, hidden node, and the number of coexisting Wi-Fi APs [26] for optimizing the ON/OFF duty cycle based on the ML method. On the other hand, hidden nodes and the number of coexisting Wi-Fi APs metrics are not so important for the LAA LBT-based coexistence scenarios. Because the LAA LBT access technique is similar to CSMA/CA of Wi-Fi, i.e the eNB must sense the availability of the medium before transmission.

Moreover, [24] proposes an adaptive LTE LBT scheme based on the Q-learning technique that is used for the autonomous selection of the appropriate combinations of Transmission Opportunity (TXOP) and muting period that can provide coexistence between co-located mMTE-U and Wi-Fi networks. Also, [27] addresses the selection of the appropriate mMTE-U configuration method based on a Conventional Neural Network (CNN) that is trained to perform the identification of LTE and Wi-Fi transmissions. In wireless resource allocation, high-quality labeled data are difficult to generate due to, e.g., inherent problem hardness and computational resource constraints [10]. Therefore, generating the dataset is one of the most important issues in the LAA/Wi-Fi coexistence scenario for training DRL models.

Although there is sufficient work without using ML on the coexistence of LTE and Wi-Fi technologies in unlicensed spectrum band, ML, and DL methods, particularly DRL based efficient channel allocation method for the densely deployed heterogeneous networks is still lacking. Moreover, there are no benchmark datasets available in densely deployed heterogeneous wireless networks for training and comparison of the ML models.

Table 1. Summary of the RRAM of communication systems with ML

Ref.	Issues addressed	Learning algorithm	Network type /Environment
Our work	Efficient channel assignment	DDQN	cellular/Wi-Fi wireless HetNets
[27]	Appropriate mMTE-U configuration	CNN	LTE/Wi-Fi HetNets
[30]	Dynamic spectrum allocation	DQN	Small BSs cellular
[31]	Dynamic multi-channel access	DDPG	LTE-WLAN HetNets
[32]	Joint allocation of spectrum, computing	DDPG	MEC-based V2X
[33]	Dynamic spectrum sharing	DQN	4G LTE and 5G NR systems
[11,24,25,26]	Spectrum sensing, spectrum allocation, and spectrum access, and channel allocation	Q-learning, DQN, DDQN, A3C	Cellular, Satellite, HomeNets and Emerging networks

Most of the DRL-based works address the problems (as mentioned above) of RRAM in cellular, satellite, HomeNets, and IoT systems instead of heterogeneous networks based on Q-learning, DQN, DDQN, and A3C algorithms. However [30-33] address the problem of heterogeneous wireless networks, such as joint optimization of bandwidth, interference management, dynamic spectrum allocation and sharing as well as power level to improve average data rate based on DRL but they have not focused on channel optimization and generating datasets in densely deployed scenarios. Table 1 summarizes these works.

### 3. Proposed Method

#### 3.1 System model

We assume an environment [14] that has a rectangular shape and consists of multiple small areas with a triangle shape as shown in figure 2.

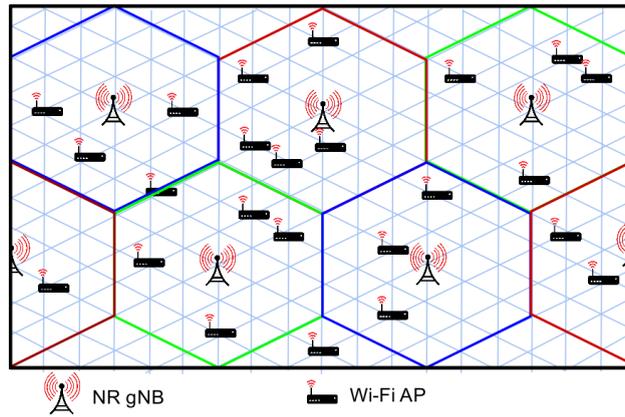


Figure 2. Assumed environment

Each small area is covered by one or more APs. The LAA BS and Wi-Fi AP coverage area was hexagonal (represented in red, blue, and green colors) and covered the same 54 triangle shapes. There are two types of users considered; Wi-Fi-only users and LTE/Wi-Fi combined users. Note that Wi-Fi-only users can use the Wi-Fi network only. The other can support the transmission and reception of both LTE and Wi-Fi traffic. Both users arrived per minimum area with an arrival rate  $\lambda$  following the Poisson arrival process. They had communication with a mean of 300 [s] following the exponential distribution and never moved until finishing their communication and the arrival ratio of each user was the same. A saturated traffic model is applied where all the nodes always have packets to transmit. As a typical scenario, we assume LAA is a cellular system in the 5GHz unlicensed spectrum band with Cat 4 LBT as a channel-sharing scheme. Here, system throughput is calculated in the case that multiple eNBs and APs share the same channel by the LBT coexistence mechanism. The LBT mechanism is modeled by the Markov chain [18] and the capacity of the LAA/Wi-Fi heterogeneous networks is calculated according to (Eqs 1 to 3).  $S_W$  and  $S_L$  represent the system throughput when LAA and Wi-Fi share the same channel by LBT, respectively.

$$S_W = \frac{P_s^W E(P_W)}{(1 - P_b)\delta + P_s^W \overline{T_s^W} + P_s^L \overline{T_s^L} + P_c^W \overline{T_c^W} + P_c^L \overline{T_c^L} + P_c^{WL} \max(\overline{T_c^W}, \overline{T_c^L})} \quad (1)$$

$$S_L = \frac{P_s^L E(P_L)}{(1 - P_b)\delta + P_s^W \overline{T_s^W} + P_s^L \overline{T_s^L} + P_c^W \overline{T_c^W} + P_c^L \overline{T_c^L} + P_c^{WL} \max(\overline{T_c^W}, \overline{T_c^L})} \quad (2)$$

$S_{W'}$  is the system throughput when Wi-Fi APs share the same channel.

$$S_{W'} = \frac{P_s^W E(P_W)}{(1 - P_b)\delta + P_s^W \overline{T_s^W} + P_c^W \overline{T_c^W}} \quad (3)$$

Let  $W$  and  $L$  denote the Wi-Fi and LAA respectively. Considered parameters for calculating system throughput are listed in table 2.

Table 2. Relevant parameters for system throughput

$E(P_W), E(P_L)$	average packet size
$P_s^W, P_s^L$	successful transmission probability
$P_c^W, P_c^L, P_c^{WL}$	collision probability
$T_s^W, T_s^L$	average time that the channel is occupied due to a successful transmission
$T_c^W, T_c^L$	average time that the channel is busy due to collision

$P_b$	channel busy probability
$\delta$	time slot

Moreover, no retry limit is considered, i.e. all the packets are ultimately successfully transmitted [18].

### 3.2. Structure of DRL based channel assignment

In this section, we propose an efficient channel assignment method for each Wi-Fi AP and cellular eNB in unlicensed bands, based on DRL. In this work, the aim of RL is to improve the decision-making ability of the central controller in wireless heterogeneous systems in the process of channel allocation so as to improve user throughput and resource utilization. Where a complex environment structure is proposed as a training environment including densely deployed APs and eNBs. We considered that eNBs are established in an environment where APs are already densely deployed. Wi-Fi APs should be managed coordinately and eNBs should have cooperated with them. For that, the agent (broker) is introduced to manage both APs and eNBs in a centralized way. Here, the state of the assumed environment is always changed due to the variation of the user's arrival and their location area information as well as the channel state in the episode. On the other hand, the learnable parameters of the agent are changing across all the episodes i.e, the agent is learning suitable actions that fit the observation state each time step. In this situation, implementing channel assignments optimally for each AP and eNB is challenging.

Therefore, we propose DRL for channel assignment to improve the user's throughput compared with other conventional methods. In brief, the optimal channel assignment provides maximum throughput for each user since it reduces channel interference and improves the capacity. Therefore in our proposed method, when training the designed Deep Q Network (DQN) agent, all possible channel assignment patterns are learned by the agent for all the explored observation states of the environment. Finally, a trained agent will be able to find an efficient channel assignment for the expected AP/eNB of the assumed environment in a short term.

Firstly the designed channel allocation problems are converted into the Markov Decision Process (MDP) framework in order to apply the DRL method [11]. In general, the aim of MDP is to define a policy to maximize the agent's reward from the environment. Therefore, we model the channel assignment problem for the proposed environment, as illustrated in figure 3, as an MDP with a state space  $S$ , action space  $A$ , transition probability  $p(S_{t+1}|S_t, A_t)$ , and reward function  $R_t(S_t, A_t)$ , where the agent is a central controller that serves as the decision maker of the corresponding action-value function. This action-value function represents the expected return after taking an action  $A_t$  in the state  $S_t$ .

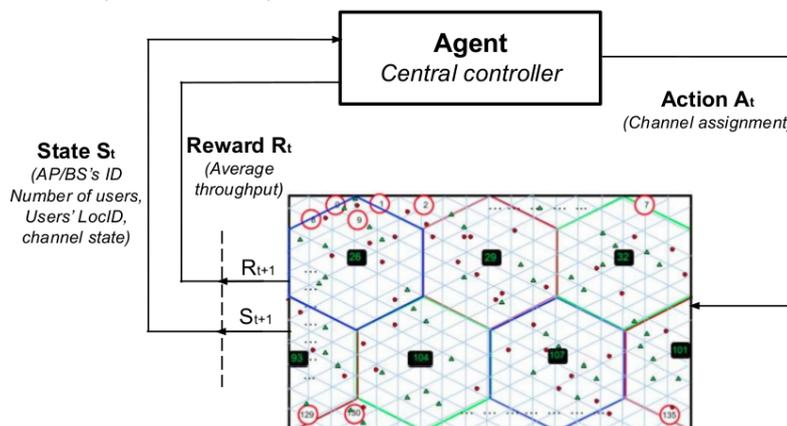


Figure 3. Interaction process between an agent and the environment

In our case for the decision-making problem, the agent/broker controls the channel to maximize the throughput by assigning suitable channels to each AP and BS in the proposed environment. In other words, the agent maps the effect of the action in a certain situation of the environment with the performed action to maximize a numerical reward signal. This mapping between the actions and rewards is called the policy rules which define the behavior of the learning agent [29]. In this environment, a random number of users connect to AP/BS in different locations for each episode. Moreover, because the broker will assign channels by avoiding the same channels to adjacent AP/BS is key to the improvement of throughput. In this research, we developed a simulator for spectrum sharing in Wi-Fi/LAA heterogeneous wireless networks based on Java as the testbed of agents. When training a DDQN agent, the average throughput is obtained from the simulator for calculating reward i.e., feedback values.

In other words, it will act as a supervisor, whose output will serve as the ground truth for training the DQN. Furthermore, when training the model, in every possible state of the environment it is learned by the agent to find optimal channel assignment patterns. Note that the agent initially has no idea about the environment. The state information is observed from our developed simulator which acts as a local server, as listed in table 3.

Table 3. State information (input data of DQN)

AP ID	Placeable area of AP/NB	Connected users	Capacity	Max capacity	User's loc area ID	Ass channel	User's Throughput
1	115	1	4.333	40	56	3	4.333
2	34	2	13.5	40	111	0	6.75
3	46	0	2.5	40	103	3	0
4	100	4	13.5	40	84	2	3.375
5	82	3	11	75	107	1	3.666
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
107	55	2	11	75	26	0	5.5

This information is used for the training Double Deep Q Networks (DDQN) agent as input data. In this case, state space is discrete, defined by four elements such as AP/BS index (placeable area the ID of AP/BS), the number of users who connected to the AP or BS, their location area index (small area ID), and assigned channel states. The state information  $S$  is preprocessed (i.e., normalize, filter, etc) before feeding to the DDQN. In other words, we filtered the state information to decrease duplication of the training data for input of DDQN, which can impact generalization performance. Since fixed information such as AP ID, AP location ID, maximum capacity, etc tend to be frequently detected in the DRL-based channel allocation problems, these duplications must be avoided.

### 3.3. Training DDQN Agent

We propose a single-agent DDQN-based DRL scheme to address the problem of efficient channel assignment in wireless heterogeneous networks. DDQN is a DQN-based method to avoid overestimations by employing two different networks, i.e.,  $Q_\theta$  and  $Q_{\theta'}$ , where  $\theta'$  is the parameter set of a target  $Q'$  network, which is duplicated from the training  $Q$  network

parameters set  $\theta$  periodically and fixed for a couple of updates.

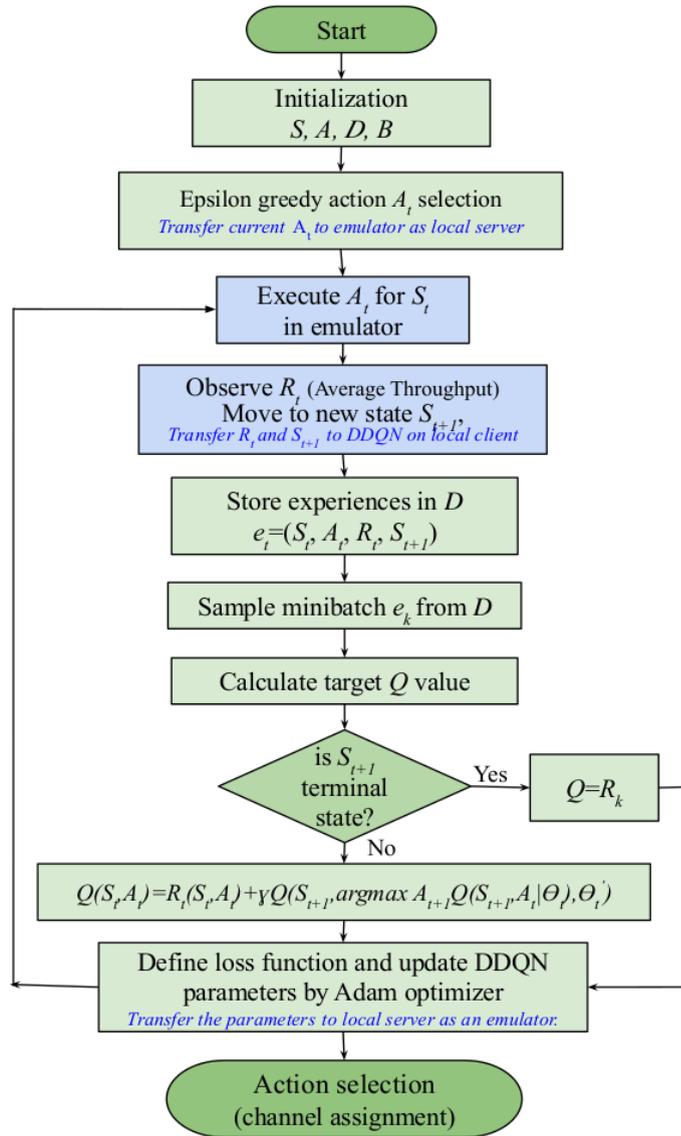


Figure 4. Flowchart of DDQN based channel assignment

**Action.** Action space is discrete, defined as the set of possible channels,  $A_t \in A = \{0, 1, 2, 3\}$ . Generally from these actions, optimal channel assignment patterns will be generated according to the epsilon greedy algorithm with  $\varepsilon=1$  random action, and  $\varepsilon=0$  greedy action for each AP and BS in the environment. In short, an optimal policy is derived from the optimal values (i.e., highest throughput) by selecting the highest-valued action in each episode.

In this work, the proposed DRL-based channel allocation scheme consists of two main parts, including a local server (environment) and a local client (DDQN agent). Between the server and client, state, action, and reward information are transferred for training our expected model. The training procedure of the proposed method is expressed in figures 4 and 5.

The input of the proposed models is the observed state from the environment where  $S_t$ , as shown in table 3 and appendix 1. The training data was extracted from our developed simulator. This dataset consists of AP ID, maximum capacity for each AP and BS, User's throughput, AP's assigned channels, user's location ID, and AP/BS's location ID information. At each time step, the agent builds its state using accumulated information from

the assumed environment. Then, the agent performs an action according to the epsilon greedy algorithm for each AP/BS in an episode. Based on this selected action and its effect on the environment, a reward function  $R_{t+1}(S_t, A_t)$  will be calculated, the higher the reward the higher the probability of choosing this performed action [29]. The output of DDQN is an expected action for channel allocation to the given AP/BS.

---

**Initialization:** experience replay memory  $D$ , state  $S_t$ , action  $A_t$ , reward  $R_t$ , experience  $e_t$  at time step  $t$  and  $Q$  with random weights

**for** all training steps **do**

Initialize and preprocess: state  $S_t$  for the new episodes

**repeat**

Select action according to Epsilon Greedy Strategy: random action when  $\varepsilon = 1$

and greedy action with  $\varepsilon = 0.01$ ,  $A_t = \operatorname{argmax}_A(Q(S_t, A, \theta))$

Transfer current  $A_t$  to emulator by TCP protocol, as a local server

Execute action  $A_t$  in a state  $S_t$  in emulator

then move to a new state  $S_{t+1}$  observe  $R_t$

Transfer  $R_t$  and  $S_{t+1}$  to DDQN on local client

Store experience:  $e_t = (S_t, A_t, R_t, S_{t+1})$  in  $D$

Sample mini batch of  $N$  transitions  $e_k$  from  $D$

Calculate target  $Q_k$  value:

**if** moved state  $S_{t+1}$  is terminal state

$Q_k = R_k$  end of the current episode

**else**

$Q_k = R_k + \gamma \max_{A_t \in A} Q(S_{t+1}, A_t)$

Define loss function:  $L = (Q_k - Q(S_t, A_t))^2$

Update neural network parameters by performing optimization algorithm Adam w.r.t actual network parameters in order to minimize the loss

Every  $C$  steps reset  $Q = Q$

**until** episode terminates (reach a certain number of iterations or when all  $Q$ -values have converged)

**end for**

---

Figure 5. Pseudo-Algorithm of DDQN for channel assessment

#### The channel assignment procedure, based on DDQN

- First, the channel state (assigned channel) information is configured as zeros for each AP/BS during the initial episodes. Note that, only one channel state of AP or BS is changed during each episode. It means that the state information in our assumed environment is able to provide  $p(S_{t+1}|S_t, A_t)$  transition probability (i.e., mapping from states in  $S$  to probabilities of selecting each action in  $A$ ) as an MDP. Additionally, the ID of AP/BS and their placeable area ID as well as maximum capacity are fixed in each AP/BS as listed in table 2. Hence the number of users who are connected to AP or BS and their location area information as well as assigned channels are assumed as random metrics in this environment.
- We defined the number of time steps for one episode as 107, relative to the number of AP/BSs deployed in the rectangular area. Where action  $A_t$  is assigned according to the epsilon greedy algorithm for each AP/BS in an episode. Then these actions (assigned

channels) are transferred to the simulator, as a local server.

- On the simulator, the reward value is calculated for assigned actions in the current state for an episode. Then this reward  $R_t$  and next state  $S_{t+1}$  information are transferred to the DDQN agent, as a local client. The differences between  $S_t$  and  $S_{t+1}$  are differentiated by the number of users, their location area information, and channel state for each time step.
- The training data (transition pairs) is stored in a replay buffer  $D$ , as  $e_t = (S_t, A_t, R_t, S_{t+1})$  at each time step. The replay memory accumulates experiences over many episodes of the MDP. When the number of  $e_t$  is reached 5000 in  $D$ , the training process will start.
- Then, DDQN updates the parameters  $Q_\theta$  and  $Q_{\theta'}$  as shown in figure 5, based on mini-batches that are constructed according to the defined batch size, as 512 from the replay buffer. The update happens only for one specific state, action pair, and for the DDQN means the loss is calculated only for one specific output unit which corresponds to a specific action. The error value of DDQN is calculated as follows:

$$Y_t^{DDQN} = R_{t+1} + \gamma Q_{\theta'}(S_{t+1}, \operatorname{argmax}_A Q_\theta(S_{t+1}, A)) \quad (4)$$

- Finally, perform the optimization according to the Adaptive Movement Estimation (Adam) algorithm with respect to actual network parameters in order to minimize this loss.
- After performing a certain number of time steps, the target network weights  $\theta'$  are updated periodically every  $C$  step according to the settings of the hyper-parameter to current network weights  $\theta$ . Repeat these steps for  $M$  number of episodes.

**Reward.** In this work, the reward function is modeled to optimize channel assignment for the assumed environment. Here, we also used a discrete reward function which provides real reward identical to average throughput, it is obtained from the assumed environment as an emulator. The process of assigning channels from a given state  $S_t$ , transitioning to a new state  $S_{t+1}$  with transition probability  $p(S_{t+1}, R_t | S_t, A_t) = P_r \{S_t = S_{t+1}, R_t = R_{t+1} | S_{t-1} = S_t, A_{t-1} = A_t\}$ .

The channel assignment of the last AP will lead to the end of an episode and the average throughput (to calculate the reward) will be reset to a new value for the new channel assignment trial. Due to the arrival rate, the location of the user, and the channel state, the target of the agent/broker changes during a channel assignment trial upon reaching a previously learned target. A DDQN agent learns these targets by simulating actions, interacting with the environment, and incurring rewards. Therefore, being able to explore new targets in an adaptive way is significant for the agents to assign the optimal channel for each AP/BS. Consequently, the trained agent is able to assign efficient channels depending on the number of users and their location (small area ID) information.

## 4. Performance Evaluation

### 4.1 Simulation model

As a simulation model, we assumed a rectangular area divided into 288 triangle areas as shown in figure 6.

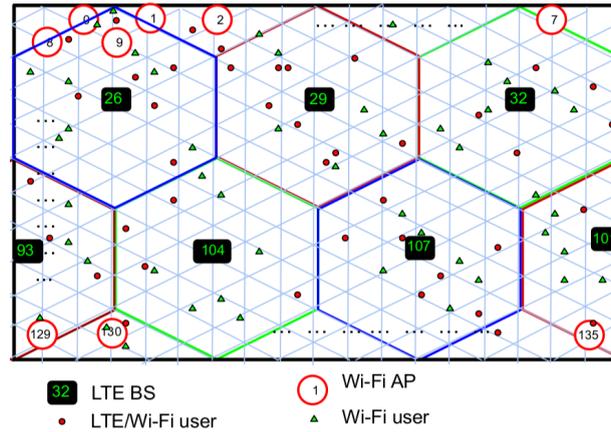


Figure 6. Simulation model

We call this triangle area, the minimum area. We assumed each cover area for LAA BS and Wi-Fi AP was a hexagonal shape (represented in red, blue, and green colors) and covered the same 54 minimum areas. The evaluation model had 136 placeable areas of LAA BSs and Wi-Fi APs. Seven LAA BSs were deployed in the center (i.e., numbered 26, 29, 32, 93, 101, 104, 107) of the hexagonal-shaped coverage area (without overlapping) as shown by black boxes. Wi-Fi APs were randomly deployed in other placeable areas. All minimum areas were covered by one or more Wi-Fi APs. The number of available channels was 4, and channels were initially assigned to Wi-Fi APs randomly. LTE BSs were assigned by using three channels so that adjacent BSs did not use the same channel and the assignments of LAA BSs were not changed during the simulations. Both Wi-Fi users and LTE/Wi-Fi combined users arrived per minimum area with an arrival rate  $\lambda$ , following the Poisson arrival process. They had communications with a mean of 300 [s] following the exponential distribution and never moved until finishing their communication. In addition, the arrival ratio of Wi-Fi users and LTE/Wi-Fi combined users were 1:1.

To evaluate the performance of NR-U/Wi-Fi heterogeneous networks in terms of average throughput, we considered an NR-U operates according to Scenario D in [28], i.e., a licensed carrier is used for uplink transmission and an unlicensed carrier is used for downlink.

Table 4. Simulation parameters of Wi-Fi and LTE for the developed simulator

Packet size	bits	12800
MAC header	bits	272
PHY header	bits	128
ACK	bits	112 + PHY header
Wi-Fi Bit Rate	Mbps	40
NR-U Bit Rate	Mbps	75
Slot Time	$\mu\text{s}$	9
SIFS	$\mu\text{s}$	16
DIFS	$\mu\text{s}$	34

Furthermore, the heterogeneous system performance is evaluated according to [18] with the parameter as shown in table 4.

## 4.2 Network architecture

When building DDQN to assign channels, we tried different settings (i.e., from minimum to the

maximum value of hyperparameters) in order to find a good hyperparameter that performs well. We trained our model according to the DDQN algorithm as shown in Figure 4,5 with the settings of hyperparameters, as listed in table 5.

Table 5. Simulation parameters of DDQN

Parameter		min value	max value	selected value
Reward	$R$	Average throughput from simulator		
Number of steps per episode	$n$			107
Number of Episodes		300	4500	2400
Update period	$Q_{\theta'}$	5	50	20
Discount rate	$\gamma$	0.9	0.999	0.99
Batch size	$e_k$	32	1024	512
Optimizer	Adam/SGD/GD			Adam
Learning rate	$\alpha$	0.0001	0.001	0.00025
Loss function	Huber/MSE/Hinge			MSE
Epsilon decay	$\varepsilon$	0.9	0.99999	0.9999
Minimum Epsilon	$\varepsilon_{min}$	0.001	0.1	0.01
Replay buffer size (max)	$D_{max}$	10000	1000000	100000
Replay buffer size (min)	$D_{min}$	500	10000	5000

From the experiment, we selected the discount  $\gamma=0.99$ , and the learning rate to  $\alpha=0.00025$ . The size of the experience replay memory was 100000. The memory was sampled to update the network every 20 steps with mini-batches of size 512. The exploration policy used was a greedy policy with the decreasing linearly from 1 to 0.01 over for each episode.

When performing the experiments, candidate channels are tried to each AP/BS step by step for each time step. Also, the reward is calculated for each selected action  $A$  in the state  $S$  for each time step. During training, the current state information of our assumed environment is given to the network's input layer as training data. Furthermore, the optimal action is selected from the actions according to the Bellman equation in the output layer which has a maximum  $Q$  value. Additionally, the experiments were performed to investigate the impact of different network architectures, optimizers, and loss functions. In our network architecture, there are two dense layers (varying the number of nodes from 8 to 288 for a layer and the number of layers from 2 to 5) as a hidden layer between the input and output layers as represented in table 6.

Table 6. List of parameters

Layer (type)	Output Shape	Total Params	Trainable params	Non-trainable params
dense (Dense)	(None, 60)	6540	8494	0
dense_1 (Dense)	(None, 30)	1830		
dense_2 (Dense)	(None, 4)	124		

The general artificial neural network model used in the experiments has three fully connected layers and there are a total 8494 of trainable parameters, 6540 for the first hidden layer, 1830 for the second hidden layer, and 124 for the output layer. All these layers are used by activation functions as Rectifier Linear Units (ReLU).

The output layers are also fully connected layers that output four actions corresponding to the predicted channel for each AP/NB according to trained DDQN agents. Thus, the output of our

DDQN for the current state  $S_t$  is  $Q(S_t) \in \mathbb{R}^4$ .

### 4.3 Simulation results

We compared the performance of the ten models which have the highest reward, from which one best model is selected. Performance of the obtained DDQN model in terms of average throughput, shown in figure 7, the horizontal axis is the number of episodes and the vertical axis is the average reward.

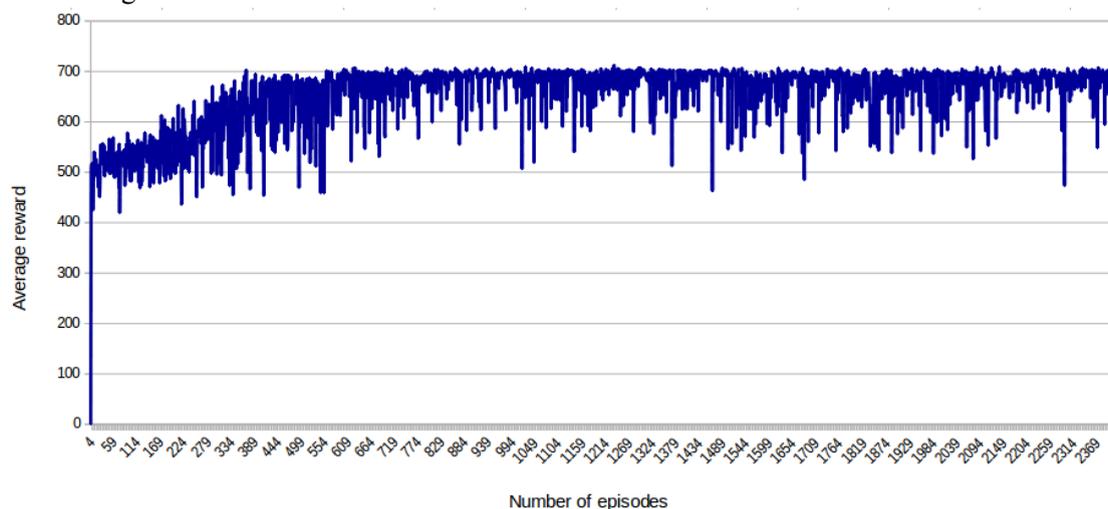


Figure 7. Performance of the obtained DDQN model in terms of average throughput

When the average reward (system throughput) converges, the agent has learned the assumed environment and can choose the optimal actions (channel assignment) in any state.

It can be observed that in about the first 100 episodes of the learning process, the average reward is almost random. This occurs as initially due to a large amount of exploration, the agent tries many different states of the assumed environment. Most of these states can not provide the desired outcome. Hence, the agent obtains small and random rewards. During the learning process, the agent locates the states that can provide optimal channel assignment for the heterogeneous network, improving the received reward. It means that the reward is converged when each user can receive the highest throughput from the available AP or eNBs in the assumed environment. After a certain number of episodes, we can observe that the learned agent can provide the desired outcome, and the average reward starts converging. The training was done over 258726-time steps. It took around 12 hours to train on 15.5 GiB of memory and Intel® Xeon(R) CPU E5-1630 v4 @ 3.70GHz  $\times$  8 of the processor. Eventually, the trained agent (broker) can assign suitable channels for each AP/gNBs in the proposed environment.

Then the training stability of our obtained model is compared with the other eight selected models, as shown in figure 8.

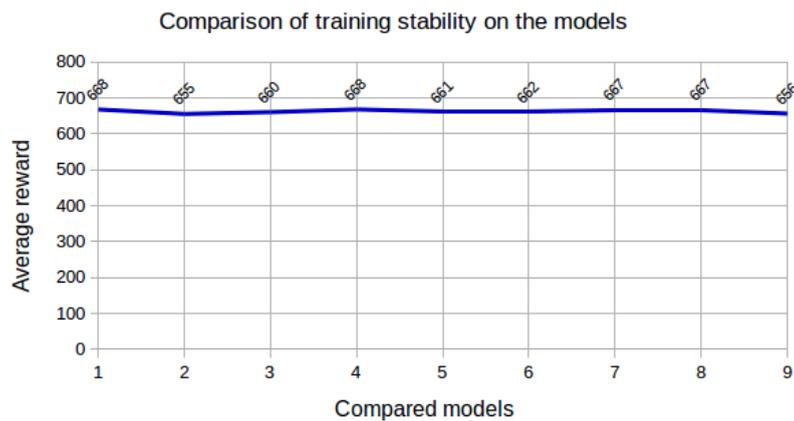


Figure 8. The average reward of compared models

These eight models are also trained on the same settings of hyperparameters and network architecture as the obtained model. Note that the first one is our selected model on the horizontal axis and the vertical axis is the average reward of the models.

The comparison of the obtained model's stability provided similar performance with the other selected models in terms of average reward (i.e., from 655 to 668). It means our obtained model can produce consistent predictions (channel assignment) concerning little changes in the environment.

We compared the coexistence performance of our proposed DRL-based channel assignment method with the random method (when disabled training section,  $\epsilon=1$ ) in the same settings of the simulator as mentioned in section 4.2. The numerical results show that our proposed DDQN algorithm improves the average throughput from 25.5% to 48.7% in different user arrival rates compared to the random channel assignment approaches. We considered five different user arrival rates, as  $\lambda=\{0.00025, 0.0005, 0.00075, 0.001, 0.00125\}$ . It means the number of users varies from 21.6 to 108 in the rectangular area for 300 msec of intervals. Therefore, when increasing the number of users in the environment, the average throughput is decreasing, as shown in figure 9.

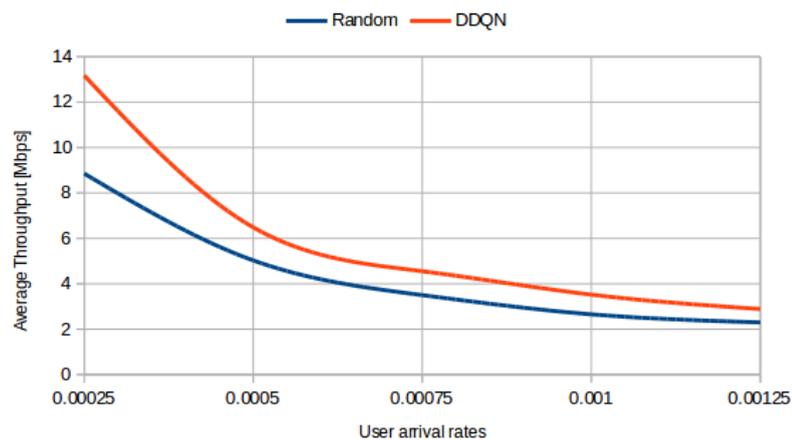


Figure 9. Comparison of average throughput in different arrival rates

We can also observe that when  $\lambda$  is less than 0.0005, the average throughput is comparatively higher than the random method.

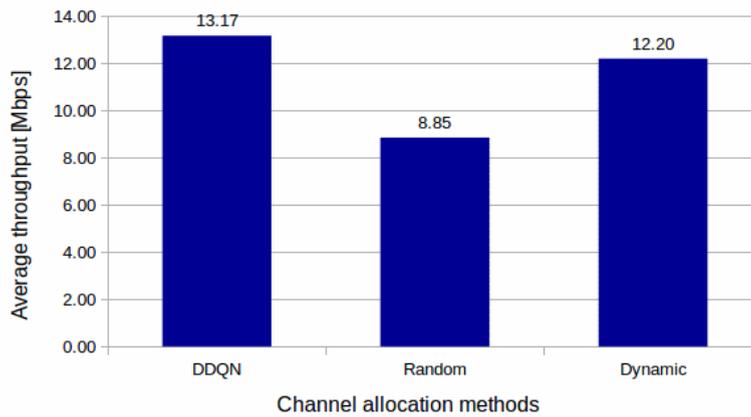
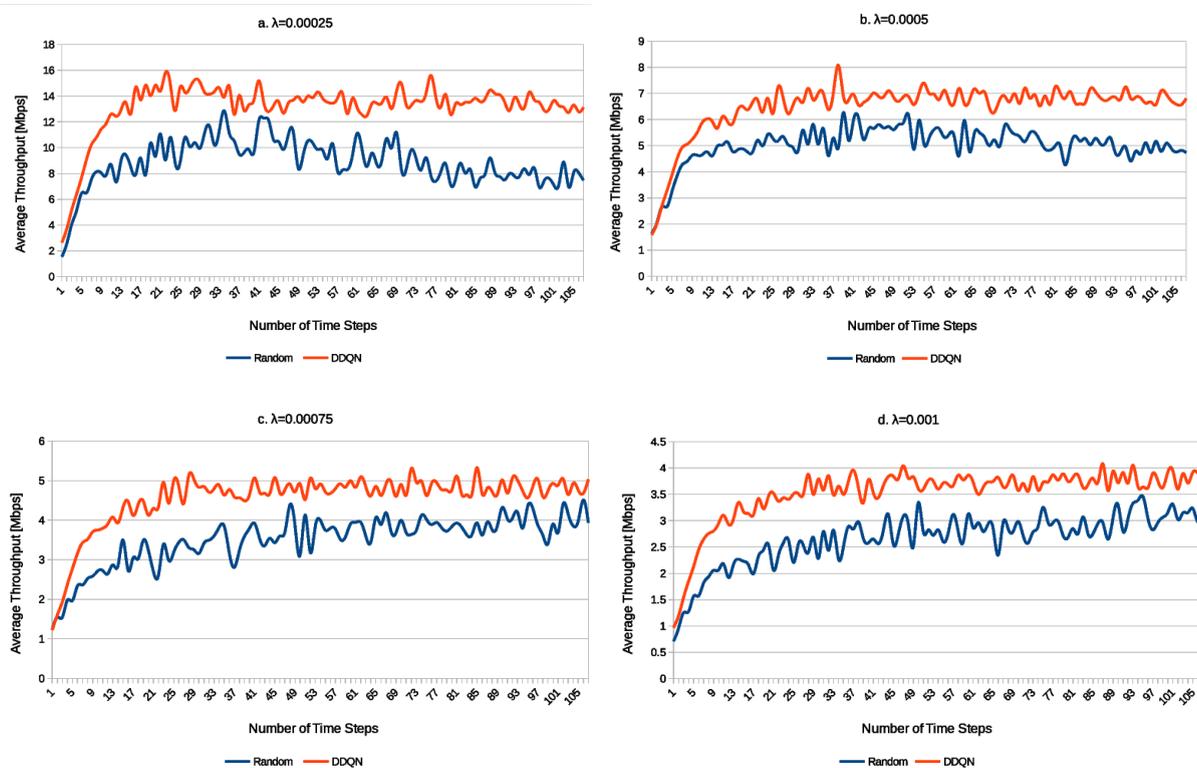


Figure 10. Comparison of the proposed method and the existing methods

In addition, we compared the average throughput of our proposed method and other existing methods, (dynamic and random) when the user arrival rate is 0.00025, as shown in figure 10. From the result, it can be observed that the proposed method can outperform the other two methods.

#### 4.4 Validation

The generalization performance of the designed DDQN has been validated for the online simulator in the same manner as the training part. We evaluated the performance of the trained agent, to confirm how well it has generalized to assign channels to the selected time steps for the episode in the proposed environment under the different user arrival rates, as represented in figure 11.



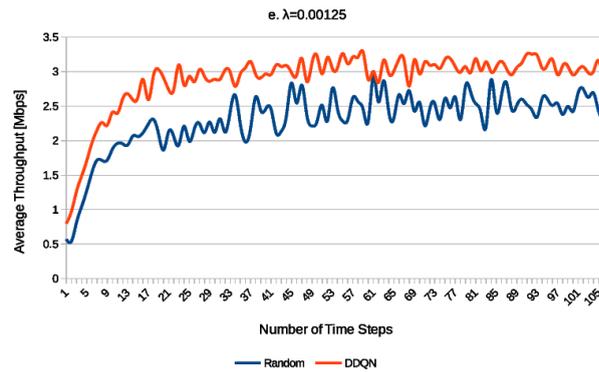


Figure 11. Validation results in different user arrival rates

Consequently, we can observe that the designed agent is trained enough to choose near optimal action with high reward for any inputs in the short term. Furthermore, we can see that from the validation result, the performance of the DDQN is impacted in terms of the user arrival rates and their location area index.

We also evaluated the stability of the obtained models (when  $\lambda=0.0005$ ) which is compared with the other eight models as mentioned in figure 8.

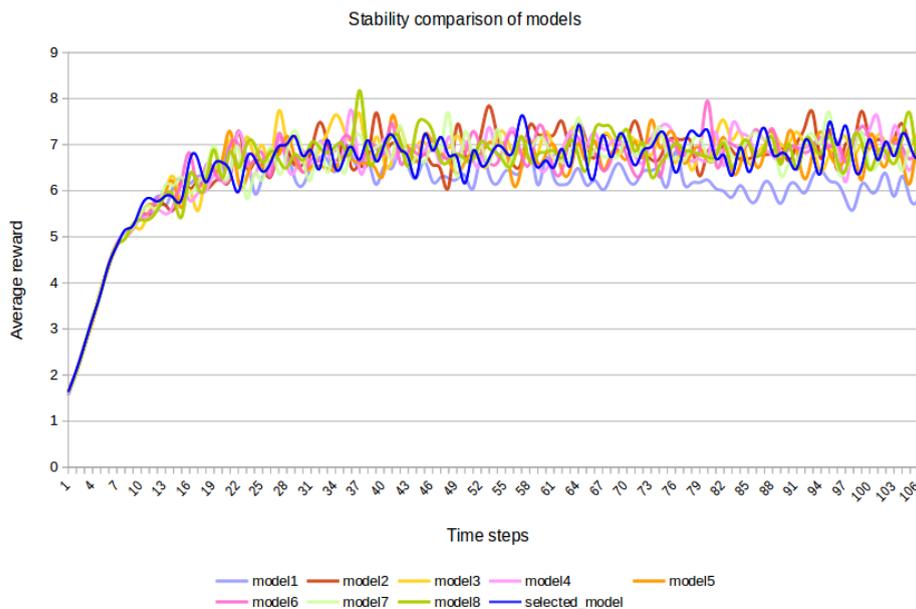


Figure 12. Comparison of stability for the obtained models

In figure 12, the horizontal axis is the number of time steps and the vertical axis is the average reward for different models. From the comparison of the models' stability, we can observe that all of the compared models provided similar performance with the selected model in terms of average reward. Moreover, the validation result of the models' stability, can provide consistent predictions for each compared model.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed to improve the average throughput in the cellular/Wi-Fi heterogeneous networks by DRL-based channel assignment method. For that, we have implemented an emulator as an environment (which was used for training models) for spectrum

sharing in densely deployed eNB and APs in wireless heterogeneous networks to train the DQN model. Additionally, based on the developed environment, the training data was generated which also can be used for training DRL-based models in an offline manner.

Regarding the obtained model, the numerical results show that our proposed DDQN algorithm improves the average throughput from 25.5% to 48.7% compared to the random channel assignment approaches. We evaluated the generalization performance of the trained agent, to confirm channel allocation efficiency in terms of average throughput (average reward) in the proposed environment under the different user arrival rates. From the evaluation results, we can observe that the trained agent can choose near optimal action with high reward for any inputs in the short term. Note that, in the performance evaluation, we assumed LTE as a cellular system since the numerical analysis of LAA throughput is available. But, the proposed method itself can be easily applied to 5G NR-U.

In the future, we will try to extend this work by modifying our environment for user mobility.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This work was supported by the Mongolia–Japan Higher Engineering Education Development (MJEED) Project (research profile code: J23A16) and JSPS KAKENHI JP20K11768.

## REFERENCES

- [1] Report ITU-R M.2370-0(07/2015) IMT traffic estimates for the years 2020 to 2030, M Series Mobile, radio determination, amateur and related satellite services.
- [2] Signals Research Group, “The prospect LTE Wi-Fi sharing unlicensed spectrum”, Qualcomm, San Diego, CA, USA, White Paper, Feb. 2015.
- [3] M. Agiwal, A. Roy, et al., "Next Generation 5G Wireless Networks: A Comprehensive Survey, *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617-1655, third quarter 2016.
- [4] F. Hu, B. Chen, et al., "Full Spectrum Sharing in Cognitive Radio Networks Toward 5G: A Survey" *IEEE Access*, vol. 6, pp. 15754-15776, 2018.
- [5] Y. Jian, C. Shih, et al., "Coexistence of Wi-Fi and LAA-LTE: Experimental evaluation, analysis, and insights", 2015 IEEE International Conference on Communication Workshop (ICCW), pp. 2325-2331.
- [6] G. Naik, J. Liu, et al., “Coexistence of Wireless Technologies in the 5 GHz Bands: A Survey of Existing Solutions and a Roadmap for Future Research”, *IEEE Communications Surveys & Tutorials*, vol 20, pp. 1777-1798, 2018.
- [7] S. Xu, Y. Li, et al., "Opportunistic Coexistence of LTE and WiFi for Future 5G System: Experimental Performance Evaluation and Analysis", *IEEE Access*, vol. 6, pp. 8725-8741, 2018.
- [8] B. Chen, J. Chen, et al., “Coexistence of LTE-LAA and Wi-Fi on 5 GHz with Corresponding Deployment Scenarios: A Survey”, *IEEE Communications Surveys & Tutorials*, 1553-877X (c) 2016.
- [9] M. Voicu, L. Simić, et al., "Survey of Spectrum Sharing for Inter-Technology Coexistence", *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1112-1144, Second quarter 2019.
- [10] L. Liang, G. Yu “Deep Learning based Wireless Resource Allocation with Application to Vehicular Networks”; arXiv:1907.03289v2, 1 Oct 2019
- [11] A. Alwarafy, M. Abdallah, et al., “Deep Reinforcement Learning for Radio Resource Allocation and Management in Next Generation Heterogeneous Wireless Networks: A Survey”; arXiv:2106.00574v1.
- [12] C. Capretti, F. Gringoli, et al., “LTE/Wi-Fi Co-existence under Scrutiny: An Empirical Study”,

October 2016 Pages 33–40, <https://doi.org/10.1145/2980159.2980164>

- [13] S. Lagen *et al.*, "New Radio Beam-Based Access to Unlicensed Spectrum: Design Challenges and Solutions", *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 8-37, First quarter 2020.
- [14] K. Kinoshita, K. Ginnan, et al., "Channel Assignment and Access System Selection in Heterogeneous Wireless Network with Unlicensed Bands", 2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 96-101, 2020.
- [15] J. Wszolek, S. Ludyga, et al., "Revisiting LTE LAA: Channel Access, QoS, and Coexistence with Wi-Fi", *IEEE Communications Magazine*, vol. 59, no. 2, pp. 91-97, February 2021.
- [16] G. Naik, J. M. Park, et al., "Next Generation Wi-Fi and 5G NR-U in the 6 GHz Bands: Opportunities and Challenges", *IEEE Access*, vol. 8, pp. 153027-153056, 2020..
- [17] M. Alhulayil, M. López-Benítez, et al., "Novel LAA Waiting and Transmission Time Configuration Methods for Improved LTE-LAA/Wi-Fi Coexistence Over Unlicensed Bands", *IEEE Access*, vol. 8, pp. 162373-162393, 2020.
- [18] Y. Gao, X. Chu, et al., "Performance Analysis of LAA and WiFi Coexistence in Unlicensed Spectrum Based on Markov Chain", 2016 *IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6.
- [19] N. Bitar, O. Al Kalaa, et al., "On the Coexistence of LTE-LAA in the Unlicensed Band: Modeling and Performance Analysis", *IEEE Access*, vol. 6, pp. 52668-52681, 2018.
- [20] Z. Tang, X. Zhou, et al., "Throughput Analysis of LAA and Wi-Fi Coexistence Network With Asynchronous Channel Access", *IEEE Access*, vol. 6, pp. 9218-9226, 2018.
- [21] C. Chen, R. Ratasuk, et al., "Downlink Performance Analysis of LTE and WiFi Coexistence in Unlicensed Bands with a Simple Listen-Before-Talk Scheme", 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), pp. 1-5, 2015.
- [22] R. Kumer, Saha al., "An Overview and Mechanism for the Coexistence of 5G NR-U (New Radio Unlicensed) in the Millimeter-Wave Spectrum for Indoor Small Cells", *Wireless Communications and Mobile Computing Volume 2021*, Article ID 8661797, 18 pages.
- [23] S. Lagen, N. Patriciello, al., "Cellular and Wi-Fi in Unlicensed Spectrum: Competition leading to Convergence", 2020 2nd 6G Wireless Summit (6G SUMMIT), pp. 1-5, 2020.
- [24] K. Nakashima, S. Kamiya, et al., "Deep Reinforcement Learning-Based Channel Allocation for Wireless LANs With Graph Convolutional Networks", Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan.
- [25] V. Maglogiannis, D. Naudts, et al., "A Q-Learning Scheme for Fair Coexistence Between LTE and Wi-Fi in Unlicensed Spectrum", *IEEE Access* PP.99.
- [26] A. Dziejdzic, V. Sathya, et al., "Machine Learning enabled Spectrum Sharing in Dense LTE-U/Wi-Fi Coexistence Scenarios"; arXiv:2003.13652.
- [27] V. Maglogiannis, A. Shahid, et al., "Enhancing the Coexistence of LTE and Wi-Fi in Unlicensed Spectrum Through Convolutional Neural Networks";
- [28] M. Hirzallah, M. Krunzet, et al., "5G New Radio Unlicensed: Challenges and Evaluation", *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 3, pp. 689-701, Sept. 2021.
- [29] N. Najem, M. Martinez, et al., "Cognitive Radio Resources Scheduling using Multi-Agent Q-Learning for LTE", *International Journal of Computer Networks & Communications (IJCNC)* Vol.14, No.2, March 2022
- [30] O. Naparstek, K. Cohen, et al., "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 310–323, 2018.
- [31] S. Wang, T. Lv, et al., "Deep reinforcement learning based dynamic multichannel access in hetnets," in 2019 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2019, pp. 1–6.
- [32] H. Peng, X. S. Shen, et al., "Deep reinforcement learning based resource management for multi-access edge computing in vehicular networks," *IEEE Transactions on Network Science and Engineering*, 2020.
- [33] U. Challita, D. Sandberg, et al., "Deep reinforcement learning for dynamic spectrum sharing of lte and nr," arXiv preprint arXiv:2102.11176, 2021.

- [34] S. Bhandari and S. Moh, "A MAC Protocol with dynamic allocation of time slots based on traffic priority in Wireless Body Area Networks", International Journal of Computer Networks & Communications (IJCNC) Vol.11, No.4, July 2019.

## APPENDIX 1

### A. Training dataset for Wi-Fi APs

Users	AP ID	Max Capacity	User Throughput	Assigned channels	User's Location ID	AP's Loc ID
0	11	40	9.00	3	66	33
1	92	40	5.20	3	256	75
2	63	40	6.75	3	123	24
3	86	40	13.50	1	185	53
4	54	40	13.50	1	261	123
5	63	40	13.50	3	107	24
6	63	40	13.50	3	35	24
7	16	40	6.50	1	227	5
8	63	40	13.50	3	17	24
9	19	40	13.00	2	44	92
10	79	40	6.50	3	147	96
11	72	40	9.00	0	259	64
12	38	40	13.50	3	17	21
13	86	40	13.50	1	187	53
14	19	40	13.00	2	44	92
15	54	40	13.50	1	242	123
16	19	40	13.00	2	64	92
17	97	40	9.00	0	179	51
18	63	40	13.50	3	106	24
19	86	40	6.75	1	259	53
20	19	40	6.50	2	5	92
21	70	40	8.67	1	39	6
22	54	40	13.50	1	243	123
23	86	40	13.50	1	187	53
24	64	40	9.00	2	2	117
25	92	40	5.20	3	221	75
184	34	40	9.00	3	183	78

### B. Training dataset for LTE BSs

Users	BS's ID	Max C	User Throughput	Assigned channels	User's Loc ID	BS's Loc ID
0	10006	75	7.83	0	226	107
1	10002	75	8.25	2	105	32
2	10001	75	13.25	1	63	29
3	10000	75	8.25	0	74	26
4	10000	75	8.25	0	2	26
5	10002	75	8.25	2	140	32
6	10005	75	9.00	2	188	104
7	10004	75	9.40	1	212	101
8	10000	75	11.00	0	91	26
9	10001	75	7.83	1	63	29
10	10000	75	33.00	0	58	26
11	10002	75	6.60	2	107	32
12	10001	75	6.71	1	63	29
13	10001	75	7.83	1	63	29
14	10001	75	11.75	1	116	29
15	10003	75	7.83	1	109	93
16	10000	75	11.00	0	130	26
17	10000	75	6.60	0	127	26
18	10003	75	6.71	1	183	93
19	10006	75	5.22	0	279	107
20	10000	75	8.25	0	21	26
21	10000	75	8.25	0	3	26
22	10004	75	9.40	1	230	101
23	10006	75	6.71	0	226	107
464	10000	75	33.00	0	5	26

## AUTHORS

**Bayarmaa Ragchaa**, She received B.S. and M.S. degrees in engineering from the Mongolian University of Science and Technology (MUST) in 2005 and 2008, respectively. She is currently a Ph.D. candidate in Information science and Intelligent systems at Tokushima University, Japan. She joined MUST as an Instructor in 2009. Her research interests include spectrum management, heterogeneous wireless systems, and mobile networks.



**Kazuhiko Kinoshita**, He received the B. E., M. E., and Ph. D degrees in information systems engineering from Osaka University, Osaka, Japan, in 1996, 1997, and 2003, respectively. From April 1998 to March 2002, he was an Assistant Professor at the Department of Information Systems Engineering, Graduate School of Engineering, Osaka University. From April 2002 to March 2008, he was an Assistant Professor at the Department of Information Networking, Graduate School of Information Science and Technology, Osaka University. From April 2008 to January 2015, he was an Associate Professor at the same University. Since February 2015, he has been a Professor at Tokushima University. His research interests include mobile networks, network management, and agent communications. Dr. Kinoshita is a senior member of IEEE and a senior member of IEICE.

