A Study on Spectrum Sharing between Cellular and Wi-Fi Networks

Doctor of Engineering

Supervisor

Professor Kazuhiko Kinoshita

Author

BAYARMAA Ragchaa

Department of Information Science and Intelligent Systems,
Graduate School of Advanced Technology and Science,
Tokushima University

March 2023

.

Table of contents

List of Figures

List of Tables

.

Abstract

In recent years, the amount of mobile traffic is growing rapidly and spectrum resources are becoming scarce in wireless networks. Under these predictions, it is clear that the wireless network capacity will not meet the exponential growth of traffic demand, such as high speed data communication, ultra reliable and low latency communication, cost effective wireless networks and so on. To overcome this problem, using cellular systems in the unlicensed spectrum has emerged as a promising and effective solution that can assist in exploiting the wireless spectrum in a more efficient way. In order to work in unlicensed bands, cellular systems, such as LTE, 5G New Radio need to coexist with legacy unlicensed technologies Wi-Fi (IEEE 802.11-based technology), Bluetooth and other systems. Consequently, providing fairness and a desired level of Quality of Service (QoS) between NR-U and Wi-Fi is a challenging issue.

In this dissertation, we propose an efficient channel assignment method for the heterogeneous wireless networks in unlicensed bands, based on Deep Reinforcement Learning (DRL) to overcome these challenges. For that, first of all we have implemented an emulator as an environment for spectrum sharing in densely deployed eNodeBs (eNBs) and Access Points (APs) in wireless heterogeneous networks to train the Double Deep Q Networks (DDQN) model. 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 cooperate with them. For that, the agent (broker) is introduced to manage both APs and NBs in a centralized way. In this case, the agent controls the channel to maximize the throughput by assigning suitable channels to each AP and BS in the proposed environment. When training the DDQN agent, the optimal channel (action) is assigned to each AP/NB based on the highest average throughput (reward) which is obtained from the emulator.

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 evaluated the generalization performance of the trained agent, to confirm channel allocation efficiency in terms of average

throughput (reward) in the proposed environment under the 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.

In the first phase of this research, we analyzed related works which includes spectrum resources, mainly unlicensed spectrum, spectrum sharing techniques as well as coexistence between the cellular and Wi-Fi systems. In Particular, the spectrum sharing between cellular LTE and Wi-Fi systems are investigated which are based on the traditional method and machine learning methods.

In the next phase of this research, we developed an environment for spectrum sharing in densely deployed eNBs and APs in wireless heterogeneous networks. Furthermore, we propose an efficient channel assignment method for each Wi-Fi AP and cellular eNB of the environment in unlicensed bands, based on the DRL. This method is aimed to improve user's average throughput compared with other existing methods. For that, a single-agent DDQN based DRL scheme is employed for efficient channel assignment problems. Consequently, our trained agent is able to assign optimal action (channels) with high reward (average throughput) depending on the number of users and their location area information.

In the final part of our work, When building DDQN, we have examined impacts of the different hyperparameter settings, different network architectures, and optimizers by experiments. The training accuracy of the designed DDQN has been validated for the on-line simulator when the training section is disabled. We evaluated the performance and the stability of trained agent, to confirm how well it has generalized to assign channels to maximum number of steps for an episode in the proposed environment under the different user arrival rates. Consequently, we can observe that the designed agent is trained enough to assign near optimal action with high reward for any inputs in the short term. Also, we can observe that from the validation result, the performance of the DDQN is impacted in terms of the user arrival rates and their location area index.

.

Chapter 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 cellular systems such as LTE and Fifth-Generation (5G) to unlicensed spectrum has emerged as a promising and effective solution that can assist in exploiting the wireless spectrum in a more efficient way and can also be a good neighbor with the other occupants [2],[3]. LTE has many advantages compared to Wi-Fi that provides the capability to carry more data traffic in a specified amount of spectrum (i.e., spectral efficiency) and which provide it an enlarged range over Wi-Fi. [2].

In principle, 5G New Radio Unlicensed (NR-U) system is allowed to operate in any unlicensed bands (from 1 to 100GHz) [4], but the initial industry focus is on the 5 GHz band. In the first phase, LTE License Assisted Access (LAA) and NR-U are expected to coexist with IEEE 802.11 based Wi-Fi systems, in the 5 GHz Unlicensed National Information Infrastructure (U-NII) bands. Also, [5] expects both LAA and NR-U to coexist in 5GHz unlicensed bands in future years.

Federal Communications Commission (FCC) opened up the U-NII radio bands at 5 GHz up to 500 MHz of spectrum that is available on a global basis for unlicensed applications [6], [7]. These bands can be classified in low frequency range (i.e., below 7 GHz) and high frequency range (i.e., ISM mmWave, around 60 GHz) frequency range for the unlicensed and shared spectrum operations, as represented figure

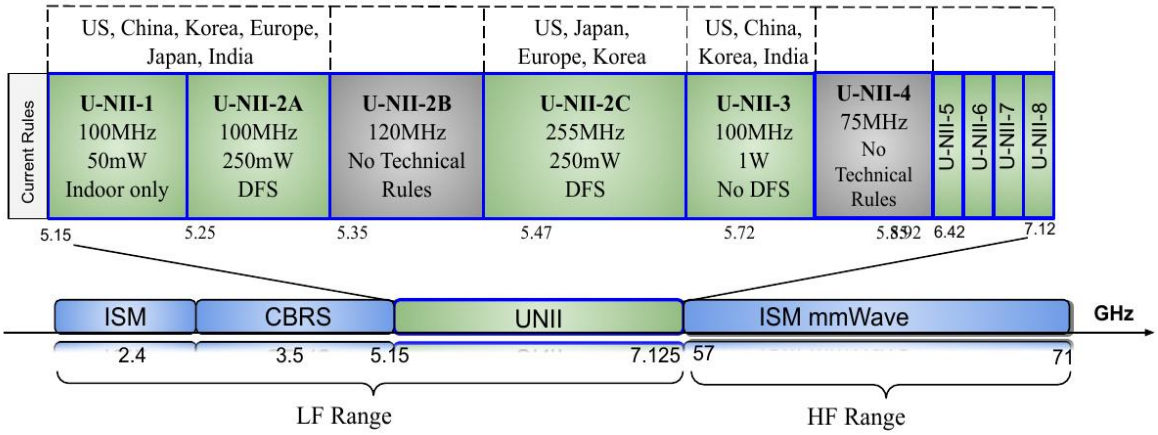


Figure 1.1: UNII spectrum bands for unlicensed and shared

1.1. Approximately 2GHz of unlicensed spectrum are applicable for the below 7GHz in omni-directional communications, which consists of Industrial Scientific Medical (ISM) band at 2.4 GHz, the Citizens Broadband Radio Service (CBRS) band at 3.5 GHz, and UNII bands at 5 GHz and 6 GHz frequencies [8]. The 5 GHz frequency bands are divided into 20 MHz bandwidth, non-overlapping channels and wider channels can be produced by bonding these primary channels. UNII spectrum bands have different constraints on their maximum transmit power, Effective Isotropic Radiated Power (EIRP), requirement for Discovery Reference Signal (DFS), applications in indoor/outdoor usage and so on. Over UNII bands, unlicensed users are mandatory to execute DFS to prevent interference with radars and other licensed operation services, whereby they have to interrupt their transmission and execute periodic sensing for radar signals.

Various heterogeneous wireless networks such as LTE LAA, 5G NR-U and IEEE 802.11 based Wi-Fi services are expected to operate in the UNII bands at 5 GHz frequency.

When different technologies operate on the same band, without any coordination, however, it causes a significant interference that reduces the average throughput per user due to the fundamental difference, as represented in table 1.1. The main difference between Wi-Fi and LAA is the contention window (CW) adjustment procedure, which appears when the contention window size CW_p achieves the maximum value CW_{max}. In Wi-Fi, once a certain number of re-transmissions have been attempted (lifetime of a packet) and if packet collision occurs, the transmitted packet

is discarded. On the other hand, LTE LAA Listen Before Talk (LBT) establishes a K parameter value in the standard. Each operator configured this value, and it takes between 1 and 8. It determines how many times the maximum value can be utilized. Once K re-transmissions have been attempted, LAA-LBT resets CW_p to CW_{min}, and re-transmission restarts from the lowest stage again. The main difference between both LTE and Wi-Fi coexistence mechanisms are listed in following table.

Table 1.1: Comparison of LTE and Wi-Fi

	LTE	**Wi-Fi**
Physical layer	OFDMA	OFDM
MAC layer	Centralized scheduling protocol	DCF protocol CSMA/CA
Bandwidth	1.4-20MHz	20MHz
Symbol duration	71.4 µs	4 µs
Modulation and coding efficiency	7.43 bits/symbol (LTE R.12)	6.67 bits/symbol (802.11ac R.12)
Retransmission	HARQ	ARQ
QoS guarantee	Yes	No
Mobility support	Yes	No

[9] 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 proposed for coexistence between LTE-U and Wi-Fi by Carrier Sensing Adaptive Transmission (CSAT) [3], [4], LBT [10], or Almost Blank Subframe (ABS) [9]. 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. [11] 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 DRL can be an effective solution and take judicious control decisions with only limited information about the network statistics [12]. There are three ways, including supervised learning, objective-oriented unsupervised learning, and reinforcement learning paradigms to incorporate Deep Learning (DL) in solving optimization problems. From these methods, we have selected to use the Deep Reinforcement Learning (DRL) approach for the efficient channel assignment problem. DRL is an advanced data-driven Artificial Intelligence (AI) technique that combines Neural Networks (NNs) with traditional Reinforcement Learning (RL). [13] investigates that DRL is a possible solution to allocate channels based on the feedback of the measured throughput considering the allocation sequence. 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 APs and cellular Base Stations (BSs). First, to apply the DRL method, the state information of the assumed environment is converted to the Markov Decision Process (MDP) framework. Therefore, we model the channel assignment problem for the proposed environment, as an MDP with a state space, action space, transition probability, and reward function, where the agent is a central controller that serves as the decision maker. The general architecture of DRL based channel allocation problem consists of two main parts, including agent and environment. We

propose a complex environment structure as a training environment included densely deployed APs and NBs. For that, we developed a simulator for spectrum sharing in Wi-Fi/LAA heterogeneous wireless networks based on Java as the testbed of agents. When training agent, the average throughput is obtained from the simulator for calculating reward. Moreover, for 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.

To validate the generalization performance of the trained agent, we employ the developed simulator in the same manner as the training section. The validation result shows that the designed agent is trained enough to choose near-optimal action with a high cumulative reward. Furthermore, it can be observed that the performance of the DDQN agent is impacted in terms of the user arrival rates and their location area index. We also evaluated the stability of the obtained model which is compared with the other eight models. This method can assist in the more efficient management of the wireless spectrum resources for the ever-increasing wireless traffic.

The rest of this thesis is organized as follows. We survey some related works, including the main spectrum sharing techniques, i.e., LBT, CSAT, and ABS. Also, spectrum sharing between Wi-Fi and cellular systems based on the traditional methods and the machine learning based spectrum sharing methods are covered in Chapter 2. Chapter 3 presents the system model of the assumed environment, the stucture of DRL based channel assignment, training procedure of DDQN agent and algorithm of

DDQN based channel assignment. Chapter 4 presents the performance evaluation of the proposed method, including simulation model, network architecture of the training model, simulation results, and validation of the proposed method and its results. Finally, Chapter 5 concludes this thesis and shows future work.

.

Chapter 2

Related Works

A cellular system operated 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 [14]. 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 [15]. A cellular (LTE/5G) system uses a centralized channel access mechanism based on Orthogonal Frequency Division Multiple Access (OFDMA).

LTE technologies that operate in the unlicensed spectrum can be divided into two categories, license-anchored systems, and non-license-anchored systems. In license-anchored unlicensed LTE systems (e.g. LTE-U, LAA), the primary carrier, referred to as the anchor, uses licensed spectrum. The anchor is used for transmissions of uplink traffic, control signaling and QoS sensitive data such as voice. The secondary carriers are used to transfer best-effort traffic and can operate in the 5 GHz unlicensed

spectrum. There are several types of cellular systems in unlicensed band, as detailed in following part:

LTE-U was developed by the LTE-U Forum in 2014 to work with the 3rd Generation Partnership Project (3GPP) Release 10-12, and coexist with WiFi using duty-cycle, where the LTE has a silence period for WiFi to have a chance to access the channel. LTE-U has been designed for operations in countries such as the US, China, and Korea, that do not mandate LBT mechanism. Further, LTE-U is supported by the ABS coexistence mechanism introduced in LTE specifications in which some (sub)frames can be unoccupied. LTE-U frames are transmitted by unlicensed spectrum bands that should be synchronized with licensed carriers in the time domain. During the OFF period of the LTE-U system, the small cell base station takes the measurements of the traffic density of neighboring Wi-Fi devices and adjusts its duty cycle appropriately. The small cell base station transmits downlink frames without performing LBT during the ON period [8]. ABSs are LTE subframes with reduced downlink transmission activity or power. Interference in pico eNBs would be less caused by macro eNBs in heterogeneous networks, by muting the transmission power of the small cell base station in certain subframes. It can be summarized that LTE-LAA activities may be controlled by an modified ABS technique in unlicensed spectrum, where uplink and/or downlink subframes may be muted, and no LTE common reference signals are involved. It is represented that Wi-Fi is allowed to reuse the blank subframes ceded by LTE, and that throughput improves with the number of null-subframes. However, since LTE throughput decreases corresponding to the number of ceded blank subframes, a tradeoff is required. Moreover, if blank subframes are non-adjacent, LTE performance reduction can be perceived since Wi-Fi transmissions are not completely restricted within LTE silent modes. During the negotiation phase, if the duration and occurrence of LTE blank subframes is reported to Wi-Fi nodes, Wi-Fi can be able to efficiently limit their transmissions within blank subframes and thus avoid interference with the LTE system. [9].

LAA was standardized by the 3GPP Rel-13 and Rel-14, and the main concept is to use carrier aggregation framework and aggregate carriers in licensed and unlicensed bands. LAA can be used as a Supplementary Downlink (SDL) or as Time Division Duplex (TDD) data channel for both uplink and downlink. LAA implements LBT protocol, which is a requirement in Europe and Japan. Before transmitting, the LAA

eNB performs Clear Channel Assessment (CCA) using energy detection [16]. LAA is modified to support scheduled uplink transmission over unlicensed spectrum bands in Rel-14, also called enhanced LAA (eLAA), as well as switching between downlink and uplink transmissions within the one channel occupancy, further enhanced LAA (feLAA).

LWA was approved as an LTE Wireless Local Area Network (WLAN) Radio Level Integration and Interworking Enhancement in 2015, and was standardized in 3GPP Release 13 in March 2016. LTE Wi-Fi Aggregation (LWA), like LTE-U and LAA, is a licensed-anchor based system which allows a mobile device to be configured by the network so that utilizes its LTE and Wi-Fi links simultaneously. However, for LTE data transmissions in the unlicensed band, LWA uses Wi-Fi based MAC and Physical Layer (PHY). LWA design primarily follows LTE DC architecture which follows a UE to connect to multiple base stations simultaneously. In the user plane, LTE WLAN are aggregated at the PDCP. Furthermore, in the control plane, eNB is responsible for LWA activation, de-activation and the decision as to which bearers are offloaded to the WLAN [17].

MulteFire Release 1.0 specification was developed by the MulteFire Alliance in 2017. It is also LTE-based technology that operates completely in the unlicensed spectrum and shared spectrum, including the global 5 GHz bands, and, does not require an anchor in the licensed spectrum [6]. Based on 3GPP Release 13 and 14, MulteFire technology supports LBT based protocol for channel access for co-existence with Wi-Fi and other technologies operating in the same spectrum. Moreover, Multefire extends its mission to support 5G private networks by developing Uni5G™ Technology Blueprints, based on current 3GPP 5G specifications, that will facilitate industries to deploy their own 5G private networks in unlicensed, shared, and locally licensed spectrum.

NR-U was developed by the 3GPP Release 16, a successor to LTE-LAA. Because 5G NR-U is developed based on the features of LAA and it supports global cellular operations in all available unlicensed spectrum bands. 5G NR-U enables operation in both Dual Connectivity (DC) mode and Carrier Aggregation (CA) mode. In DC mode, a User Equipment (UE) is able to exchange data with multiple base stations at the same time, where one base station is designed as the master base station and the

others are as secondary base stations. On the other hand, a UE exchanges data with a single base station through two or more contiguous or non-contiguous component carriers that could be in-band or out-band along the CA mode. The case of in-band CA, both primary and secondary carriers are placed within the same band, whereas for the out-band CA, the carriers can be placed in different spectrum bands. DC mode improves both throughput and reliability, but it is more complex and expensive compared to CA mode. On the other hand, CA mode improves the throughput only. 3GPP provides the five different adaptable NR-U deployment scenarios based on DC or CA which is used to connect with UEs in the unlicensed carriers, as follows:

- Scenario A: Carrier aggregation between licensed band NR Primary Cell (PCell) and NR-U Secondary Cell (SCell), a UE is supported by a licensed carrier through a 5G NR cell and an unlicensed carrier through an NR-U cell. Where, NR-U SCell may have both downlink and uplink, or downlink-only. NR PCell is connected to 5G Core Network (CN).

- Scenario B: Dual connectivity between licensed band LTE PCell and NR-U PSCell, a UE is supported by a licensed carrier through a LTE primary cell and an unlicensed carrier through an NR-U cell. LTE PCell connected to Evolved Packet Core (EPC) has higher priority than PCell connected to 5G-CN.

- Scenario C: Stand-alone NR-U, A UE is supported primarily by a NR-U cell. NR-U is connected to 5G-CN. This scenario is suitable for operating private networks.

- Scenario D: A stand-alone NR cell in unlicensed band, A UE is supported by a licensed carrier through an NR cell for uplink communication, and by an unlicensed carrier through an NR-U cell for downlink communications. NR-U is connected to 5G-CN.

- Scenario E: Dual connectivity between licensed band NR and NR-U, a UE is supported by a licensed carrier through an NR cell and an unlicensed carrier through an NR-U cell. PCell is connected to 5G-CN

A critical difference between NR-U and previous 3GPP-based unlicensed RATs is that NR-U does not require a licensed primary carrier for its operation [18]. Moreover, the systems differentiate in their timing resolution, number of possible uplink and

downlink occurrences for a Channel Occupancy Time (COT), as well as their Hybrid Automatic Repeat Request (HARQ) specifications.

CSMA/CA mechanism: In contrast, Wi-Fi employs OFDM digital modulation scheme and the DCF as a fundamental access mechanism to the wireless medium, which is structured to be asynchronous and decentralized. The CSMA/CA mechanism is used as a channel access method for Wi-Fi systems. Before a transmission, Wi-Fi has to sense the channel for a fixed period of time Arbitration Inter-Frame Space (AIFS), (i.e., defer duration), in LAA/NR-U. This procedure is called CCA. Figure 2.1 illustrates the CSMA/CA procedure. Only if the channel is available for DCF Interframe Space (DIFS) duration, the node able to start transmission.

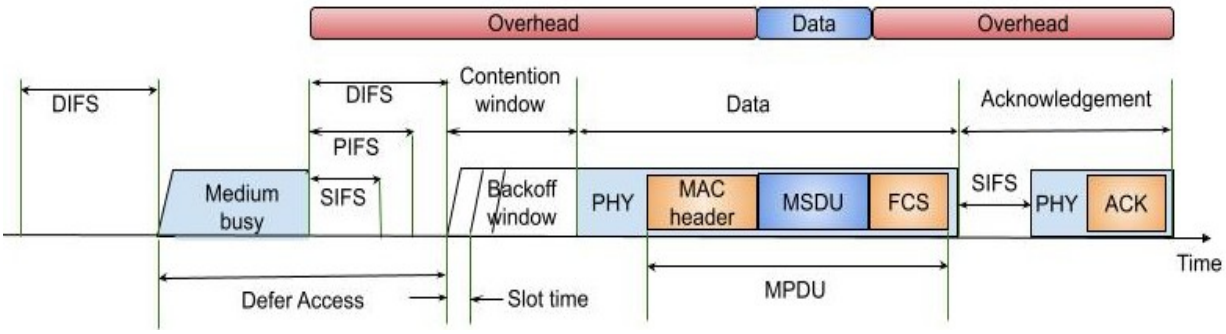


Figure 2.1: CSMA/CA protocol

Furthermore, prior to a new transmission immediately after a successful transmission, the node has to postpone its transmission for DIFS with the addition of a random backoff time. The back-off slots indicate how many idle time slots a node has to sense before a transmission. The number of the timeslots is determined by the backoff counter that is randomly chosen from a uniform distribution over the interval $[0, CW - 1]$. When the transmission is not successful an Acknowledgment (ACK) signal is not received. Then the node arranges a retransmission after a new exponential backoff period until the maximum number of retransmissions is achieved. At each unsuccessful transmission, CW is doubled, up to a maximum value of contention window $CW_{max} = 2^m CW_{min}$.

Carrier Sense (CS) and Energy Detection (ED) functions are included in the CCA procedure. The CS function involves the capability of the receiver to detect and decode a received Wi-Fi preamble. On the other hand, when the receiver is not able

to decode the received signal, ED function is employed. The threshold of CS and ED is -82 and -62 dBm for 20 MHz bandwidth respectively, as specified by the standards [19].

LBT-based channel access mechanism: 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 [10].

To address this issue, a coexistence mechanism is required to manage the interference between two different technologies. To this end, a number of 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 [20]. The markets including Europe, Japan, and India that require regulation in the unlicensed spectrum need more robust equipment to periodically check for the presence of other occupants in the channel (listen) before transmitting (talk) on millisecond scale.

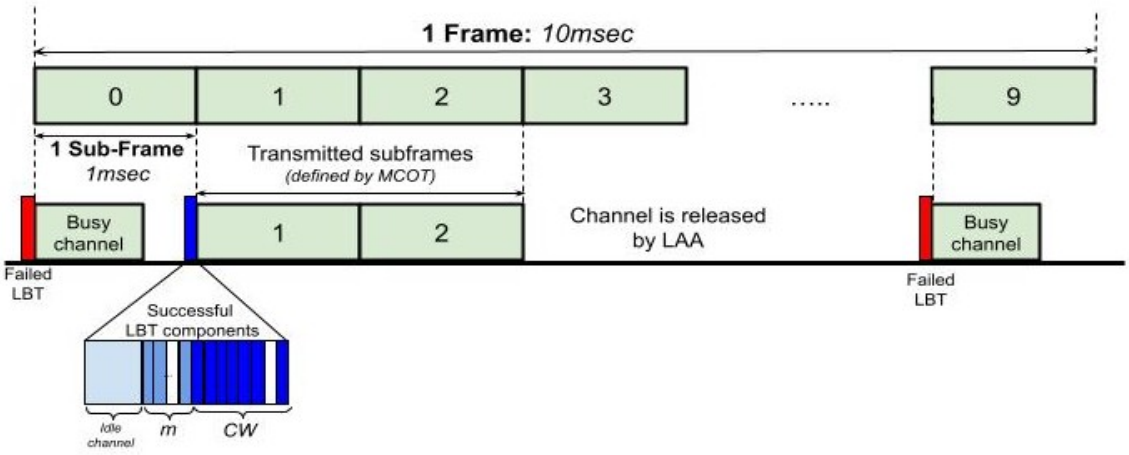


Figure 2.2: The basic LBT-based channel access mechanism

The main principle of LBT (as represented by the blue color, in figure 2.2) is defined as follows:

- A transmitter before starting a transmission, first waits for the channel to be idle for 16mms.

- The device performs CCA after each of the 'm' required observation slots.
- For the back off-stage, the device selects a random integer N in 0, ..., 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 acknowledgement (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 [21].

Two types of LBT mechanisms are employed in LTE-LAA mandated by European Telecommunications Standards Institute (ETSI). One is Frame Based Equipment (FBE) (figure 2.3) and other Load based Equipment (LBE) (figure 2.4).

FBE-Based LBT Mechanism: In this mechanism, equipment are permitted to perform CCA to sense if the channel is idle, and this is settled for every fixed frame period. When the current operating channel becomes idle, the equipment immediately can transmit for a duration equivalent to the channel occupancy time (COT) [22]. Where, unlicensed equipment contends for the channel beginning only at synchronized frame boundaries. Furthermore, if the operating channel becomes busy (i.e., occupied by other users), the equipment is unable to transmit for the next fixed frame period on that channel. The transmission time is fixed and it varies between minimum 1ms and maximum 10ms. Therefore, if the equipment has an opportunity of channel access, it occupies it for a fixed time period, COT specified by the operator, and then waits for a period equal to 5% of COT, for the next transmission. The FBE-based LBT mechanism is simple for the design of reservation signals and requires less standardization.

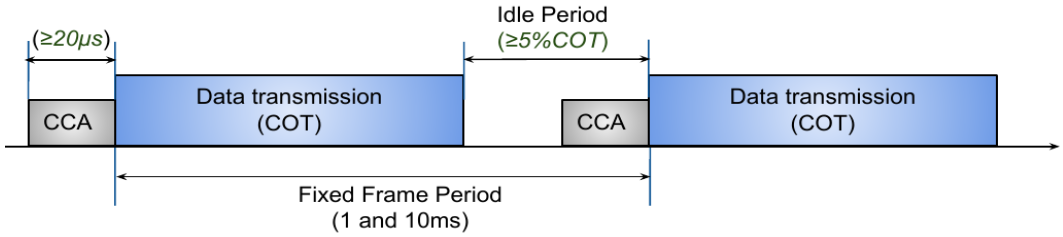


Figure 2.3: FBE-Based LBT Mechanism

LBE-Based LBT Mechanism: It is another channel access mechanism based on LBT, in this method, the equipment is required to specify whether the channel is free or not. Unlike FBE, LBE is based on the demand and not dependent on a fixed frame period. In the case where the unlicensed equipment detects a free operating channel, it will immediately start transmission. If there is no free channel, an Extended Clear Channel Assessment (ECCA) is performed, where the channel is detected for a period of random factor N multiplied by the CCA time slot. N is the amount of free slots so that a total idle period should be observed before transmission.

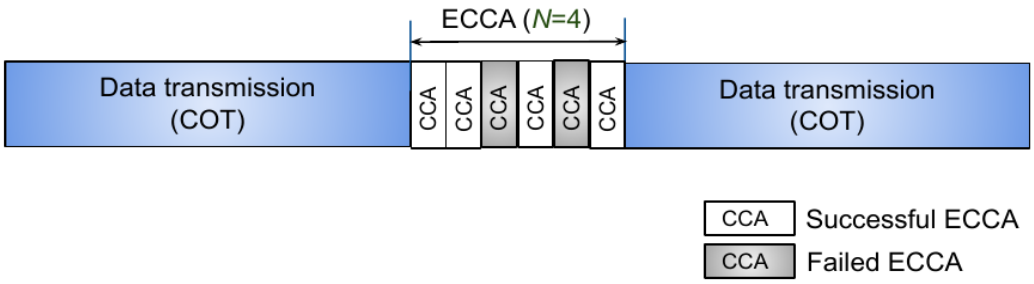


Figure 2.4: LBE-Based LBT Mechanism

Its value is selected randomly from 1 to q, where q contains a value between 4 and 32. The counter will decrease by one when a CCA slot is idle. Once the counter reaches zero, the equipment is able to start transmission. Moreover, the maximum channel occupancy time is determined by $(13/32) \times q\, ms$. Therefore, the maximum channel occupancy time is 13 $[ms]$ when q equals to 32 which is the best coexistence parameter [23].

Channel access categories: To facilitate LTE-LAA and NR-U operation over unlicensed bands, four LBT Categories (CATs) have been defined:

- CAT1-LBT: An NR-U device accesses the channel directly without performing LBT.
- CAT2-LBT: An NR-U device senses the channel for a fixed time duration, T_{fixed}. If the channel remains idle during this period, the device can access the channel.
- CAT3-LBT: An NR-U device backs off for a random period of time before accessing the channel. This random period is sampled from a fixed-size contention window.
- CAT4-LBT: An NR-U device backs off for a random period of time before accessing the channel, similar to the CSMA/CA procedure with exponential backoff.

The base station can perform CAT2-LBT procedure before sending critical messages, e.g DRS, which are crucial for initial access and network detection. The DRS frame contains basic information for supporting initial access in unlicensed bands, i.e., Synchronization Signal Block (SSB), in the 5G NR system. The SSB contains the Physical Broadcast Channel (PDCH) and synchronization signals, i.e., Primary Synchronization Signal (PSS) and Secondary Synchronization Signal (SSS). The CAT4-LBT is investigated as the main procedure for the channel access in unlicensed spectrum bands. [8].

Channel Access Priority Classes: For LAA operation, 3GPP adopts the CAT4-LBT scheme, which is similar to EDCA but considers different parameters for accessing the unlicensed spectrum. LAA defines four priority classes. The deferment period T_{df} in LAA is equivalent to AIFS in Wi-Fi. The airtime in LAA is referred to as COT, and the maximum COT (MCOT) for different priority classes. During the MCOT period, the SBS sends an OFDMA frame, where it schedules resource blocks (distributed across time and frequency) to user equipment. In LAA, SBS infers the failure of transmission by monitoring the HARQ-ACK feedback messages sent by UEs over the licensed channel. LAA supports smaller AIFS values and hence LAA devices are expected to capture channels faster than those with Wi-Fi, resulting in an unfair situation. During a COT, multiple DL and UL occasions can be initiated in which UEs are assigned to different resources that are distributed in time, frequency, and spatial domains. The CAT2-LBT is required if the time to switch between DL and

UL exceeds a certain limit, i.e., 16 microseconds [8].
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 in order to improve system performance.

2.2 Spectrum sharing between Wi-Fi and cellular without ML

The several survey and tutorial papers [6], [9], [24], [18] 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, [10] 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.

Using this framework, they present a literature review on intertechnology coexistence with a focus on wireless technologies with equal spectrum access rights, i.e., primary/primary, secondary/secondary, and technologies operating in a spectrum commons. Moreover, the possible spectrum sharing design solutions and performance evaluation approaches useful for future coexistence cases are identified in this work.

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. In addition, they present a comprehensive taxonomy of spectrum sharing in Cognitive Radio (CR) networks from the perspective of Wider-Coverage, Massive-Capacity, Massive-Connectivity, and Low-Latency four application scenarios. Particularly, the key enabling technologies that may be closely related to the study of 5G in the near future are summarized in terms of full-duplex spectrum sensing, spectrum-database based spectrum sensing, auction based spectrum allocation, carrier aggregation based spectrum access.

[6] addresses coexistence issues between a number of important wireless technologies such as LTE/Wi-Fi as well as radar operating in the 5GHz bands, with a particu-

lar focus on four coexistence scenarios. Also, the research provides brief descriptions of wireless technologies such as Wi-Fi, LTE, radar, and Dedicated Short-Range Communications (DSRC) operating in the 5 GHz bands. [9] 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. Focusing on those important issues, this paper surveys the coexistence of LTE-LAA and Wi-Fi on 5 GHz with corresponding deployment scenarios, and introduces a scenario-oriented decision-making method for coexistence. [20] 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. Authors evaluate their results by comparing with traditional channel assignment methods by simulation experiments. Their proposed method enhances both the average user throughput and the fairness of user throughputs compared with conventional static channel assignment methods. [21] evaluates the impact of LAA under its various QoS settings on Wi-Fi performance in an experimental testbed. Especially, they considered several issues such as, clarifying LBT rules, including a description of the changes introduced in the latest ETSI and 3GPP standards and methods for ensuring QoS, evaluating the impact of LAA under its various QoS settings on Wi-Fi performance in a standardized experimental testbed as well as identifying research challenges for 3GPP technologies in unlicensed bands.

Various methods proposed in [25] to adapt the transmission and waiting times for LAA based on the activity statistics of the existing WiFi network which is exploited to tune the boundaries of the CW. Moreover, a dynamic method is proposed to adapt the TxOP times for LAA based on the HARQ feedback. The methods are evaluated using the ns-3 network simulator based on the 3GPP fairness definition. The results show that selecting fixed waiting times for LAA based on the existing Wi-Fi activities is more friendly to the existing Wi-Fi and provides better total aggregated throughputs for both coexisting networks compared to the 3GPP Cat 4 LBT algorithm. Furthermore, the proposed dynamic TXOP method is more friendly to the existing Wi-Fi and provides better total aggregated throughputs compared to the fixed TxOP period approach of the 3GPP Cat 4 LBT scheme. [14] provide the LTE and Wi-Fi behavior when sharing the same spectrum while operating under a broad range of network conditions. Specifically, they deploy a test bed with commodity Wi-

Fi hardware and low-cost software-defined radio equipment running an open-source LTE stack. The user-level performance attainable over LTE/Wi-Fi technologies is investigated when employing different settings, including LTE duty cycling patterns, Wi-Fi offered loads, transmit power levels, modulation and coding schemes (MCS), and packet sizes. The obtained results demonstrate that duty cycling patterns are key to the throughput performance attainable by Wi-Fi, but also impact on the jitter performance important to real-time applications, under homogeneous power settings LTE can lock out Wi-Fi transmissions, if not alternating silent/active periods, as transmit power is increased, Wi-Fi load negatively impacts on LTE throughput, no single LTE transmission strategy ensures Wi-Fi performance is maximized when operating with different MCSs and packet sizes, and Wi-Fi contention levels do not affect LTE performance. Their results show that optimizing the performance of both technologies requires not-easy tuning of several parameters while closely monitoring Wi-Fi operation and application-specific requirements.

The interference impact of LAA-LTE on Wi-Fi is studied in [5] under various network conditions using experimental analysis in an indoor environment. In this paper propose the problems that are likely to arise due to the coexistence of LAA-LTE and Wi-Fi in indoor environments using experimental evaluation. The critical PHY layer parameters and design are investigated in five experiments that explore how LAA-LTE interference impacts Wi-Fi performance, as follows.

- Wi-Fi throughput can be heavily degraded by LAA-LTE transmissions with 3/ 5/ 10MHz bandwidth (especially 3/5MHz)
- LAA-LTE transmissions can have a small impact on Wi-Fi throughput when using a 1.4 MHz channel with center frequencies located on the guard bands or the center frequencies of Wi-Fi channels.
- LAA-LTE transmissions with 1.4/3/5MHz bandwidth can trigger Wi-Fi CS/CCA and thus heavily impact Wi-Fi performance.
- Wi-Fi with MIMO can perform worse than Wi-Fi without MIMO when LAA-LTE interference is strong.
- Increasing distance between LAA-LTE and Wi-Fi links does not necessarily decrease the impact of interference in the indoor environment. On the other

hand, blocking Line Of Sight (LOS) between LAA-LTE and Wi-Fi links can effectively help decrease the impact of interference.

Based on these experimental results, the design of the MAC protocol can be guided. [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 [26]-[27] to evaluate the downlink performance of coexisting LAA and Wi-Fi networks by using the Markov chain. Particularly, [22] 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 [18]. Therefore, initially, NR-U is expected to coexist with Wi-Fi and LTE-LAA technologies in the 5GHz unlicensed spectrum band. [28] 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, [29] 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. All studies come to the same conclusion, namely that coexistence mechanisms are required to enable coexistence between co-located LTE and Wi-Fi networks [30].

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 [12], [13], [31], [32] are proposed for the communication system problem, especially for RRAM optimization problems such as channel and spectrum allocation and spectrum access, etc. RRAM plays a pivotal role during infrastructure planning, implementation, and resource optimization of modern wireless networks. Efficient RRAM solutions will provide enhanced network connectivity, improved system efficiency, and reduced energy consumption [12]. Particularly, [13] 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 mechanism in LTE-U/Wi-Fi heterogeneous networks mostly focus on the power control, hidden node, and the number of coexisting Wi-Fi APs [32] 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, [13] proposes an adaptive LTE LBT scheme based on the Q-learning technique that is used for autonomous selection of the appropriate combinations of TXOP and muting period that can provide coexistence between co-located mLTE-U and Wi-Fi networks. Also, [33] addresses the selection of the appropriate mLTE-U configuration method based on a 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 [11]. Therefore, generating the dataset is one of the most important issues in the LAA/Wi-Fi coexistence scenario for training DRL models.

[34] addresses a dynamic multichannel access problems based on DQN, where multiple correlated channels track an unknown joint Markov model. A user at each time slot determines a channel to transmit data and receives a reward based on the success or failure of the transmission to maximize cumulative rewards. Moreover, they provide an analytical study on the optimal policy for fixed-pattern channel switching with known system dynamics and show through simulations that DQN can achieve the same optimal performance without knowing the system statistics. 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 are still lacking. Moreover, there are no benchmark datasets available in densely deployed heterogeneous wireless networks for training and comparison of the ML models.

In [35], a multi-agent DQN-based model that jointly tackles the dynamic channel selection and interference management in SBSs cellular networks that share a set of unlicensed channels in LTE networks. In the proposed scheme, the SBSs are the

agents who choose one of the available channels for transmitting packets in each time slot. The agent's action is channel access and channel selection probability. The DQL input includes the channels' traffic history of both the SBSs and WLAN, while the output is the agent's predicted action vectors. Simulation results reveal that their proposed DQL strategy enhances the average data rate by up to 28% when compared to the conventional Q-learning scheme.

In [36], a single-agent DQN-based model is proposed to tackle the dynamic spectrum allocation for multiple users that share a set of K channels in the same network.

[37] address a single-agent prediction Deep Deterministic Policy Gradient (DDPG) based DRL algorithm to examine the problem of the dynamic Multi Channel Access (MCA) for the hybrid LTE-WLAN aggregation in dynamic HetNets. The agent is the central BS controller, whose state space is continuous, consisting of both the channels' service rates and the users' requirement rates. The action space, on the other hand, is discrete, representing the users' index. Two reward functions are provided; online traffic real reward and online traffic prediction reward, each of which are functions of users' requirements, channels' supplies, degree of system fluctuation, the relative resource utilization, and the quality of user experience. Using simulation results, the authors demonstrate the efficiency of the proposed prediction-DDPG model in solving the dynamic MCA problem compared to conventional methods.

[38] consider the joint allocation of the spectrum, computing, and storing resources in Multi-access Edge Computing (MEC) based vehicular networks. In particular, the authors propose multi-agent DDPG-based DRL algorithms to address the problem in a hierarchical fashion considering a network composed of Macro eNodeB (MeNB) and Wi-Fi APs. The agents are the controller installed at MEC servers. The agents' action space is discrete including the spectrum slicing ratio set, spectrum allocation fraction sets for the MeNB and for each Wi-Fi AP, computing resource allocation fraction, and storing resource allocation fraction. The state space is discrete, representing information of the vehicles within the coverage area of the MEC server, including vehicles' number, x-y coordinates, moving state, position, and task information. The reward function is discrete, defined in terms of the delay requirement, and requested storing resources required to guarantee the QoS demands of an offloaded task. Provided experimental results reveal that their proposed schemes achieve high QoS satisfaction ratios compared with the random assignment techniques.

[30] propose a single-agent DQN algorithm based on Monte Carlo Tree Search

(MCTS) to address the problem of dynamic spectrum sharing between 4G LTE and 5G NR systems. In particular, they used the MuZero algorithm to enable a proactive BW split between 4G LTE and 5G NR.
The agent is a controller located at the network core, whose action space is discrete, corresponding to a horizontal line splitting the BW to both 4G LTE and 5G NR. The state space is discrete, defined by five elements, including an indicator if the user is an NR user or not, the number of bits in the user's buffer, an indicator of whether the user is configured with Multimedia Broadcast Single Frequency Network (MBSFN) or not, the number of bits that can be transmitted for the user in a given subframe, and the number of bits that will arrive for each user in the upcoming subframes.

Table 2.1: 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 mLTE-U configuration	CNN	LTE/Wi-Fi HetNets
[32]	Dynamic spectrum allocation	DQN	Small BSs cellular
[33]	Dynamic multi-channel access	DDPG	LTE-WLAN HetNets
[34]	Joint allocation of spectrum, computing	DDPG	MEC-based V2X
[35]	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

The reward function is a continuous function explained as a summation of the exponential of the delayed packet per user. Experimental results indicate that their proposed method provides comparable performance to the state-of-the-art optimal solutions. Most of the DRL based works address the problems (as mentioned above) of RRAM in cellular, satellite, HomNets and Internet of Things (IoT) systems instead

of heterogeneous networks based on Q-learning, DQN, DDQN and Asynchronous Advantage Actor Critic (A3C) algorithms.

However [30]-[36] 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 to focus on channel optimization and generating datasets in densely deployed scenarios. Table 2.1 summarizes these works.

.

Chapter 3

Proposed Method

3.1 System model

We assume an environment [20] that has a rectangular shape and consists of multiple small areas with a triangle shape as shown in figure 3.1.

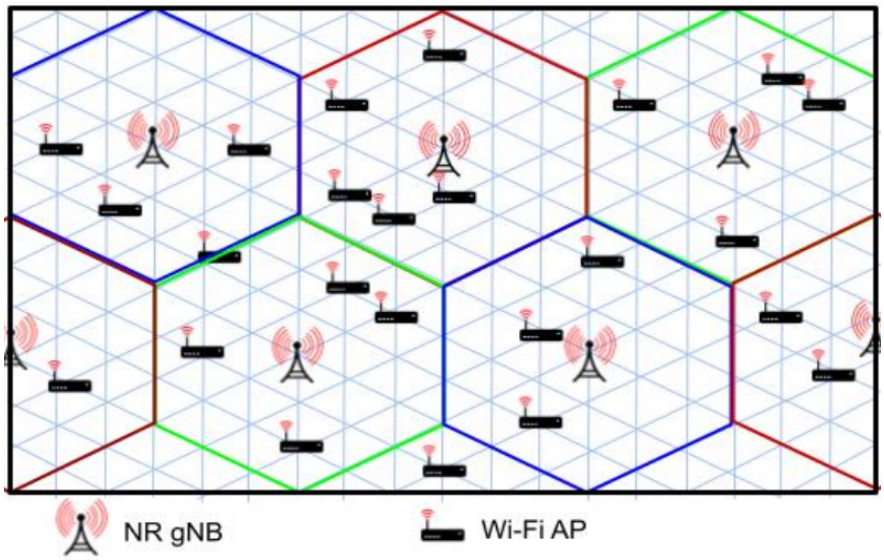


Figure 3.1: Assumed environment

In this assumed environment, Wi-Fi APs are deployed randomly, and LTE BSs are deployed so that their coverage area is not overlapped with other BSs. Moreover, Wi-Fi APs cannot avoid channel interference, and LTE BSs cause interference to Wi-Fi APs since LTE BSs also use unlicensed bands. The whole coverage area can be

served by one or more Wi-Fi APs, and a user can connect to the AP wherever. 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. In this assumed environment, 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 λ following the Poisson arrival process. These users can be covered by one or more coverage areas of AP/BS. But at the moment of time, an user is available to connect to only one AP or BS. They had communications 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 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.

Channel Access Probability with Cat 4 LBT LAA. With Cat 4 LBT scheme, if there is a new transmission buffered at an idle LAA eNB, it executes CCA to detect the availability of an unlicensed carrier. If the channel is detected to be free, the LAA eNB can transmit immediately. If CCA become unsuccessful, LAA-LBT launches extended-CCA (ECCA) stage 0, with CW of 16. The ECCA stage increments by one, and the CW size doubles (until the maximum ECCA stage of 6 and the maximum CW size of 1024, respectively) every time an unsuccessful transmission happens. If a packet transmission of an eNB become unseccessful, when reaching the maximum ECCA stage, the ECCA stage and CW size will reset to their initial values.

The counter value is an integer randomly selected from the CW size of ECCA stage $m(0, CW_m - 1)$. The counter is decremented by one if the channel is idle for a time slot, and freezes when the channel becomes busy. The eNB starts transmission when the counter reaches zero. The eNB enters an idle state after a successful transmission. The state of an LAA eNB is demonstrated by a stochastic process ($s(t)$, $z(t)$), where (-1, 0) indicates the state after a successful CCA for this Cat 4 LBT LAA mechanism. Moreover, $s(t) \in (0, 1...m-1, m)$ indicates the ECCA stage, $z(t)$ means the counter value and $CW_{s(t)} = CW_{min}2^{s(t)}$ means the CW size in stage $s(t)$.

Under unified transmission failure probability p_f, the channel busy probability p_b, and packet arrival rate q, the Cat 4 LBT mechanism is modeled as a Markov chain according to [26], and the capacity of the LAA/Wi-Fi heterogeneous networks is calculated by (Eqs 1 to 3). Where authors considered the same carrier sense threshold for both Wi-Fi and LAA, a free-space propagation channel, and that all the nodes in the coexistence scenario can detect the other nodes' signal above the carrier sense threshold.

System performance. The system throughput can be calculated as follows [26].

$$S = \frac{E[P]P_s}{E[T]} \tag{3.1}$$

S_W and S_L represent the system throughput when LAA and Wi-Fi share the same channel by LBT, respectively.

$$S_W = \frac{P_s^W E(\bar{P}_W)}{(1-P_b)\delta + P_s^W T_s^{\bar{W}} + P_s^L T_s^{\bar{L}} + P_c^W T_c^{\bar{W}} + P_c^L T_c^{\bar{L}} + P_c^{WL} max(T_c^{\bar{W}}, T_c^{\bar{L}})} \tag{3.2}$$

$$S_L = \frac{P_s^L E(\bar{P}_L)}{(1-P_b)\delta + P_s^W T_s^{\bar{W}} + P_s^L T_s^{\bar{L}} + P_c^W T_c^{\bar{W}} + P_c^L T_c^{\bar{L}} + P_c^{WL} max(T_c^{\bar{W}}, T_c^{\bar{L}})} \tag{3.3}$$

Furthermore, the throughputs are calculated not only when LTE and Wi-Fi share the channel but also when Wi-Fis share the same channel. $S_{W'}$ is the system throughput when Wi-Fi APs share the same channel.

$$S_{W'} = \frac{P_s^W E(\bar{P}_W)}{(1-P_b)\delta + P_s^W T_s^{\bar{W}} + P_c^W T_c^{\bar{W}}} \tag{3.4}$$

Let W and L denote the Wi-Fi and LAA respectively. Considered parameters for calculating system throughput are listed in table 3.1.

$$T_s^W = \frac{H + E(P)}{R_W} + \delta + SIFS + \frac{ACK}{R_W} + DIFS + \delta \tag{3.5}$$

$$T_c^W = \frac{H + E(P)}{R_W} + \delta + DIFS + \frac{ACK}{R_W} + DIFS + \delta \tag{3.6}$$

T_s^L, T_c^L are defined as follows:

Table 3.1: Relevant parameters for system throughput.

$E(P_W), E(P_L)$	average packet size
$E[T]$	average length of a time slot
P_s^W, P_s^L	successful transmission probability
$P_c^W, P_c^L, P_c^W L$	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
δ	time slot
H	MAC and PHY header size
ACK	acknowledge frame size
R_W, R_L	bit rates
C_W, C_L	capacity of Wi-Fi AP and LTE BS
$u_{Wn'}, u_{Ln'}$	number of users connected to AP or BS

$$T_s^W = \frac{H + E(P)}{R_L} + \delta + \frac{ACK}{R_L} + DIFS + \delta \tag{3.7}$$

$$T_c^L = \frac{H + E(P)}{R_L} + \delta + DIFS + \frac{ACK}{R_L} + DIFS + \delta \tag{3.8}$$

When n_W Wi-Fi APs and n_L LTE BSs are mixed, the system throughput is calculated by Eqs (3.2) and (3.3). Consequently, the capacity per AP or BS are defined as follows,

$$C_W = \frac{S_W}{n_W} \tag{3.9}$$

$$C_L = \frac{S_L}{n_L} \tag{3.10}$$

Consequently, a Wi-Fi user and a LTE+Wi-Fi user can obtain throughput as follows,

$$\frac{C_{W_n'}}{u_{W_n'}}[Mbps], \tag{3.11}$$

$$\frac{C_{L'_n}}{u_{L'_n}}[Mbps] \tag{3.12}$$

respectively. Moreover, no retry limit is considered, i.e. all the packets are ultimately successfully transmitted [26].

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 with DQN based DRL scheme. 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. In particular, many Wi-Fi APs are deployed uncoordinatedly and contend to employ spectrum resources, so that the users' obtained throughputs are degraded seriously. To overcome that, Wi-Fi APs should be managed coordinately and eNBs should have cooperated with them. It will improve the efficiency of spectrum usage and the quality of communication for users. 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 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.

The Markov Decision Process. Firstly the designed channel allocation problems are converted into the MDP framework in order to apply the DRL method [12]. It provides a mathematical framework for modeling decision-making problems whose outcome is random and controlled by an agent. For studying optimization problems, MDPs are useful that can be solved by dynamic programming and reinforcement learning techniques. In a reinforcement learning procedure, an agent can learn its optimal policy via interaction with its environment by trial and error to maximize the long-term reward. In particular, the agent first observes its current state, then takes an action, and receives its immediate reward together with its new state [23]. The MDP is typically represented mathematically by the tuple (S, A, p, R). In general, the aim of MDP is to define a policy to maximize the agent's the cumulative reward $\pi^* = max_\pi R$ from the environment. Therefore, we model the channel assignment problem for the proposed environment, as illustrated in figure 3.2, 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.

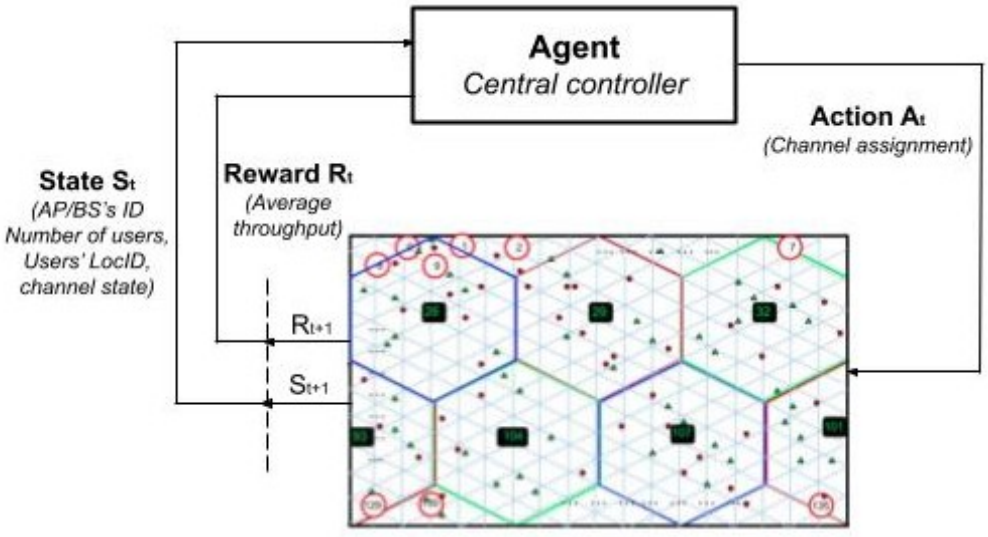


Figure 3.2: Interaction process between an agent and the environment

This action-value function represents the expected return after taking an action

A_t in state S_t. This function is essential as they illustrate the relationship between the MDP mathematical formulation and the DRL formulation [12]. A_t time t, the agent observes a state S_t from the state space S. The state space should contain useful and effective information about the wireless heterogeneous network environment. Then, the agent selects action A_t from the action space A, such as the channel allocation. The selected action must provide the desired result, such as average throughput maximization. Then the state S_t moves to a new state S_{t+1} with a transition probability p, and the agent receives a feedback numerical reward R_t which evaluates the quality of the taken action.

This interaction, i.e., (S_t, A_t, R_t, S_{t+1}), between the agent and the wireless environment repeatedly continues, and the agent will utilize the received instantaneous reward to adjust its strategy until it learns the optimal policy π^*. The agent's policy π defines the mapping from states to the corresponding actions $S \leftarrow A$, i.e., $A_t = \pi(S_t)$. DRL algorithms to handle MDP problems belong to two major groups of approaches; which are the value-based and the policy-based methods. as shown in table 3.2.

Table 3.2: DRL algorithms used in RRAM

Family	**Algorithm**	**Action space**	**Policy type**
Value based	Q-Learning	Discrete (discrete state space)	Off
	DQN	Discrete	
	Dueling DQN	Discrete	
	Double DQN	Discrete	
Policy based	Reinforce	Discrete and continuous	On
	A2C-A3C	Discrete and continuous	On
	DDPG	Continuous	Off

Value-Based Algorithms. This group of approaches is applied to evaluate the value function of the agent. This value function is then utilized to implicitly and greedily obtain the optimal policy. There are two types of value functions, such as the value function and the state-action function. Both define the expected, accumulated, discounted rewards received when taking action A_t in state S_t for the value function. If not at pair (S_t, A_t) for the state-action function and then following the policy π thereafter.

These functions are very important as they describe the connection between the mathematical formulation of MDP and the DRL formulation, which are specified as follows [23].

$$V^{\pi}(s) = E\left[\sum_{t=0}^{\infty} \gamma^{t} r_t(s_t, a_t, s_{t+1}) | a_t \sim \pi(.|s_t), s_0 = s\right] \tag{3.13}$$

$$Q^{\pi}(s, a) = E\left[\sum_{t=0}^{\infty} \gamma^{t} r_t(s_t, a_t, s_{t+1}) | a_t \sim \pi(.|s_t), s_0 = s, a_0 = a\right] \tag{3.14}$$

The optimal value function $V^*(S)$ and state-action function $Q(S, A)$ is approximated by the Bellman's optimality equations, as follows,

$$V^*(s) = max_{a_t}\left[r_t(s_t, a_t) + \gamma E_{\pi} V^*(s_{t+1})\right] \tag{3.15}$$

$$Q^*(s, a) = r_t(s_t, a_t) + \gamma E_{\pi}\left[max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1})\right] \tag{3.16}$$

The main purpose of MDP is to acquire the optimal policy π^* i.e., mapping states to optimum actions with the highest reward. Thus, the best action could be found to be the ones that maximize the above value functions, and the optimal policy will be the one that maximizes these values functions [12], [23]. Specifically, the Q-function is $Q^{\pi}(S, A)$ is utilized for finding the optimal policy, $\pi = argmax Q^{\pi^*}(s_t, a_t)$. The foremost aim of the value based DRL algorithm is to approximate this function.

Q-Learning method. In RL, Q-learning is one of the most common algorithms to manage MDPs. It obtains the optimal values of the Q-function iteratively utilizing the updating rule of the Bellman equation, as follows,

$$\begin{aligned} Q(s_t, a_t) = Q(s_t, a_t) + \alpha_t[r_t(s_t, a_t) + \\ + \gamma max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \end{aligned} \tag{3.17}$$

Where α_t t is the learning rate that determines how much the new information promotes the existing Q-value. The key idea of this Bellman rule requires discovering the TD between the current Q-value $Q(s_t, a_t)$ and the expected Q-value,

$$r_t(s_t, a_t) + \gamma max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$$

This optimal policy demonstrates that in any state, an agent takes the action that will ultimately receive the highest cumulative reward. However, the Q-learning algorithms have many constraints when applied to RRAM in modern wireless networks. First, it is applicable only to issues with low dimensionality of both state and action spaces, pushing it unscalable. Next, it is applied only to RRAM with discrete state space and action space, such as channel selection and Radio Access Network (RANs) assignment. But, there are many real-world applications of reinforcement learning that require an agent to select actions from continuous spaces. When used for the problem with continuous action space, such as autonomous controls, and power allocation, the action space must be discretized. Consequently, results will be inaccurate because of the quantization error.

Deep Q Network. Even though the Q-learning algorithm is based on making a table for the Q values, it becomes unsuccessful to obtain the optimal policy when the state space and action space become relatively large. This issue frequently happens in the RRAM problems of modern wireless systems. To overcome that, the DQN algorithm was developed, which acquired the advantages of Q-learning and DL approaches. The key idea is to replace the table in the Q-learning algorithms with a Deep Neural Network (DNN) that attempts to approximate the Q values. Therefore, the DNN has also named the function approximator and indicated it as $Q(S_t, A_t|\theta)$, where θ demonstrates the training parameters (i.e., weights) of the DNN. The replay memory is demonstrated by D, and it is generally used for interrupting the relationship between the training samples and transitions, i.e., (S_t, A_t, R_t, S_{t+1}), by making them independently and identically distributed i.i.d. The replay memory stores the training transitions during the learning process of the policy, that are generated with the interaction process of the wireless environment. The DQN's agent will then randomly select minibatch samples of transitions from D to train its DNN. For improving the stability of the DQN model, the target Q network is used, whose weights will be periodically updated to track those of the main Q network.

Since the DQN algorithm is mostly used to learn the optimal policy, i.e., $\pi^* = argmaxQ^{\pi^*}(S_t, A_t)$, the optimal Q-function is obtained from the iterative Bellman equation, as follows:

$$Q(s_t, a_t) = r_t(s_t, a_t) + \gamma max_{a_{t+1}} Q(s_{t+1}, a_t) \tag{3.18}$$

Then the DQN algorithm is optimized by iteratively updating the training parameters, θ of its DNN to minimize the Bellman loss function, as follows;

$$\begin{aligned} L(\theta_t) = E_{s_t, a_t, r_t, s_{t+1}} \in D[r_t(s_t, a_t) + \\ + \gamma max_{a_{t+1}} Q(s_{t+1}, a_t|\theta') - Q(s_t, a_t|\theta)]^2 \end{aligned} \tag{3.19}$$

Where θ' is the training parameter of the target Q network. The DQN technique can be utilized for channel allocation, access control, spectrum access, user association, and RANs assignment efficiently. DQN algorithm can also be used for RRAM problems with continuous action space, such as power control, by discretizing the action space. However, such a methodology makes DQN ineffective to critical quantization errors that may considerably decrease its accuracy. There are also other constraints in the basic DQN, and different Q-learning algorithms are addressed to overcome that.

Double DQN. In the DQN algorithm, the overestimation error may appear for the Q values, which can degrade the training efficiency and lead to suboptimal policies. The overestimation error results from the positive bias caused by the max operation utilized in the Bellman equation. In particular, the core reason is that the same training transitions are used for selecting and evaluating an action. Several methods were proposed for this problem to use two Q value functions, one for selecting the best actions and the other to evaluate the best actions. The action selection process is still according to the online weights parameters θ, while the second weights parameters θ' are used to evaluate the value of this policy. Hence, the worth of the policy is still estimated based on the current Q values, as in conventional Q learning. The weights parameters θ are updated via swapping between θ and θ' parameters. Therefore, the target Q values are produced from the modified Bellman equations, as follows:

$$Q(s_t, a_t) = r_t(s_t, a_t) + \gamma Q(s_{t+1}, argmax_{a_{t+1}} Q(s_{t+1}, a_t|\theta_t), \theta'_t) \tag{3.20}$$

and the Double DQN algorithm applies the modified Bellman loss function to update its weights, as follows;

$$\begin{aligned} L(\theta_t) = E_{s_t, a_t, r_t, s_{t+1}} \in D[r_t(s_t, a_t) + \\ + \gamma Q(s_{t+1}, argmax_{a_{t+1}} Q(s_{t+1}, a_t|\theta_t), \theta'_t) - Q(s_t, a_t|\theta_t)]^2 \end{aligned} \tag{3.21}$$

The Double DQN algorithm is also broadly used in RRAM problems, due to its advantages compared to the basic DQN algorithm.

Problem definition. 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 consequence of the action in a particular condition of the environment with the performed action in order to maximize a numerical reward signal. This mapping between the actions and rewards is called the policy rules that describe the behavior of the learning agent [16]. 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 3.3.

On the developed simulator, the simulation period was divided into some time slots with 300 seconds of constant length. At the end of each time slot, extract the number of users in each small area, the assigned channel for each AP, and the throughput of users who had finished their communication during the time slot. This

Table 3.3: 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

information is used for the training DDQN agent as input data. This case, state space is discrete, defined by four elements such as AP/BS index (placeable area ID of AP/BS), 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. We can choose the most appropriate DRL algorithm depending on the dimensionality of the RRAM problem, that fits the problem settings. For example, RRAM problems could have discrete action space, such as channel allocation, channel access, and RAN assignment, etc. Therefore, we selected a value-based DDQN algorithm because of the discrete action space. DDQN is a DQN based method to avoid overestimations by employing two different networks, i.e., Q_θ and $Q_{\theta'}$, where θ' is the training parameter set of a

target Q network, which is duplicated from the training Q network parameters set Q_θ periodically and fixed for a couple of updates. The problem of overestimation is that the agent always chooses the non-optimal action in any given state because it has the maximum Q value. When such a problem occurs, the noises from estimated Q value will cause large positive biases in the updating procedure. As a consequence, the learning process will be very complicated and messy.

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 $e = 1$ random action, and $e = 0$ 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 components, such as a local server (environment) and a local client (DDQN agent). Between the server and client, a state, action, and reward information is transferred by Transmission Control Protocol (TCP) for training our expected model. The training procedure of the proposed method is depicted in figures 3.3 and 3.4. Note that an action selection in state and a reward calculation process are implemented in the local server (emulator), as represented in the blue color in figure 3.3. The other procedures are executed in the local client, i.e., training the proposed DDQN agent.

The input of the proposed models is the observed state from the environment where S_t, as indicated in table 3.3 and the appendix C,D. The training data was extracted from our developed simulator. This dataset comprises 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 $R_{t+1}(S_t, A)$ will be calculated, the higher the reward the higher the probability of choosing this performed action [39]. The output of DDQN is an expected action for channel allocation to the given AP/BS.

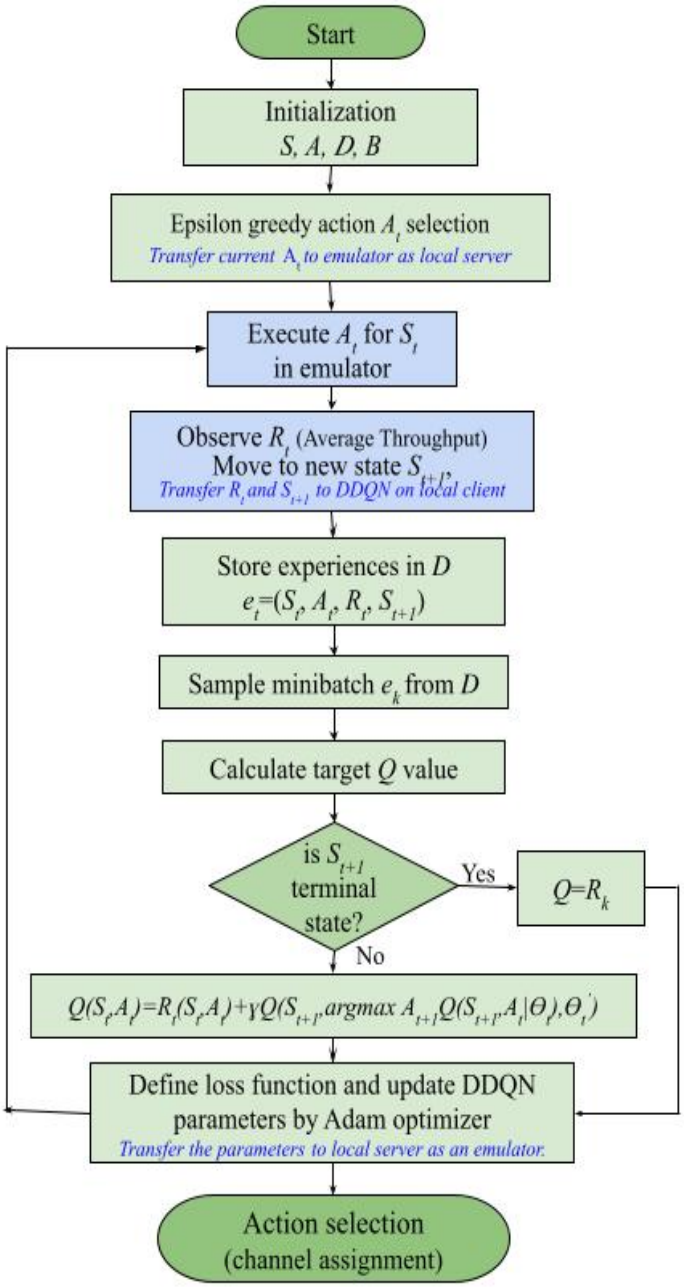


Figure 3.3: Flowchart of DDQN based channel assignment

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

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_1 for the new episodes

repeat

Select action according to Epsilon Greedy Strategy: random action with ε =1
and greedy action with ε=0.01, $A_t = argmax_A(Q(S_t, A, \theta))$

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 3.4: Pseudo-Algorithm of DDQN for channel assessment

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 a 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 3.3. 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 and 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 by TCP protocol.

- 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 states S_t and S_{t+1} are differentiated by the number of users, their location area information, and channel state for each time step.

- During the learning process of the policy, the training transitions as $e_t = (S_t, A_t, R_t, S_{t+1})$ are stored in replay buffer D, that are generated during the interaction with the wireless environment. 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_θ and $Q_{\theta'}$ as shown in figure 3.3, 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 that 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:

$$Y_t^{DDQN} = R_{t+1} + \gamma Q_{\theta'} - (S_{t+1}, argmax_A Q_\theta(S_{t+1}, A)) \tag{3.22}$$

- Finally, perform the optimization according to the 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 θ' are updated periodically every C step according to the settings of the hyper-

parameter to current network weights θ. 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 in the environment 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, trained agent is able to assign efficient channels depending on the number of users and their location (small area ID) information.

.

Chapter 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 4.1.

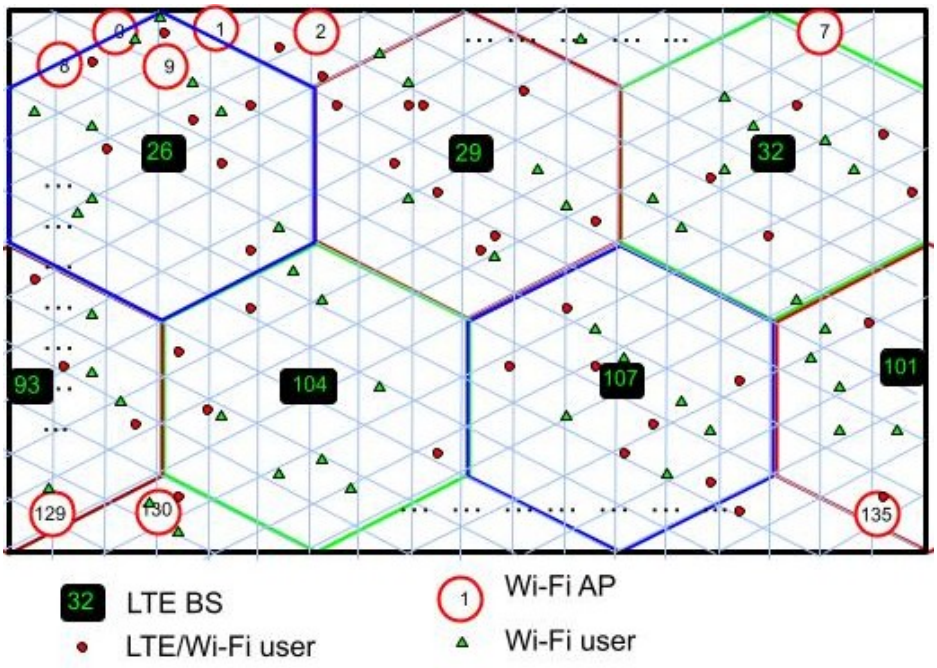


Figure 4.1: Simulation model

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.

There are two types of users considered as mentioned 3.1, both Wi-Fi users and LTE/Wi-Fi combined users arriving per minimum area with an arrival rate λ, 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.

Table 4.1: Simulation parameters of Wi-Fi and LTE

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	μs	9
SIFS	μs	16
DIFS	μs	34

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 [8], i.e., a licensed carrier is used for uplink transmission as a primary carrier and an unlicensed carrier (secondary carrier) is used for downlink. The main difference between the two carriers is that the primary carrier is also responsible for communicating most signaling and control information, including system acquisition, authentication, mo-

bility management, access, paging, registration, and control information associated with the secondary carrier. The secondary carrier in the unlicensed spectrum will be more opportunistic and only used in a way so that it shares the spectrum fairly with other systems that are using the spectrum, including Wi-Fi. Furthermore, the heterogeneous system performance is evaluated according to [26] with the parameter as shown in table 4.1.

4.2 Network architecture

When building DDQN to assign channels, we tried different settings (i.e., from minimum to 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 3.3 and 3.4, with the settings of hyperparameters, as listed in table 4.2.

Table 4.2: Simulation parameters of DDQN

Parameter		**min value**	**max value**	**selected value**
Reward	R	A	verage 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	γ	0.9	0.999	**0.99**
Batch size	e_k	32	1024	**512**
Optimizer	Adam/SGD/GD			**Adam**
Learning rate	α	0.0001	0.001	**0.00025**
Loss function	Huber/MSE/Hinge			**Huber**
Epsilon decay	ϵ	0.9	0.99999	**0.9999**
Minimum Epsilon	ϵ_{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 $\gamma = 0.99$ which is applied to the future rewards. The learning rate was $\alpha = 0.00025$ and 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.

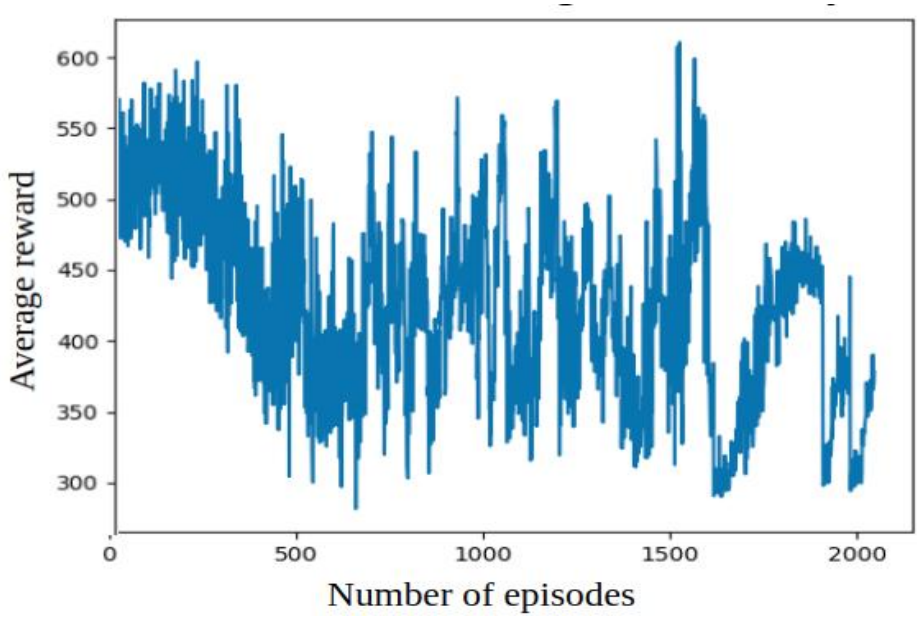


Figure 4.2: When number of nodes is high in each layers

Additionally, the experiments were performed to investigate the impact of different network architectures, optimizers, and loss functions. When we increased the number of nodes and layers, the performance of the model was decreasing and it was not generalizing, as shown in figure 4.2.

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 4.3.

Table 4.3: 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. Our selected network architecture represented in figure 4.3. During training, the current state information of our assumed environment is given to the network's input layer as training data.

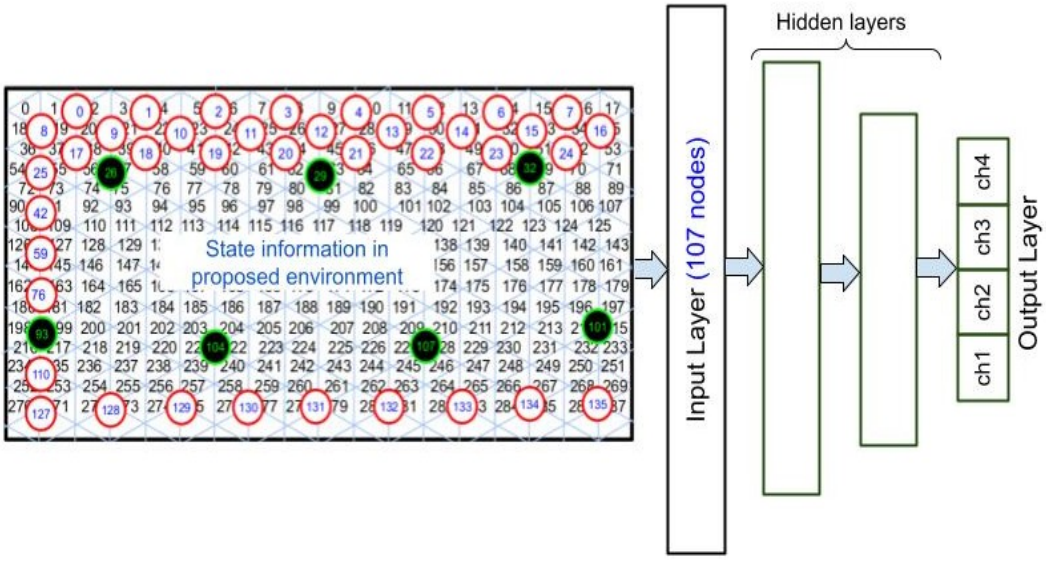


Figure 4.3: Selected network architecture

All these layers are used by activation functions as ReLu. Moreover, we observed that the number of hidden layers is more than two for the network architecture, it

was reducing the performance of the proposed agent. As well as, we tried dropout technique which also reduced the performance of our proposed method, as shown in figure 4.4.

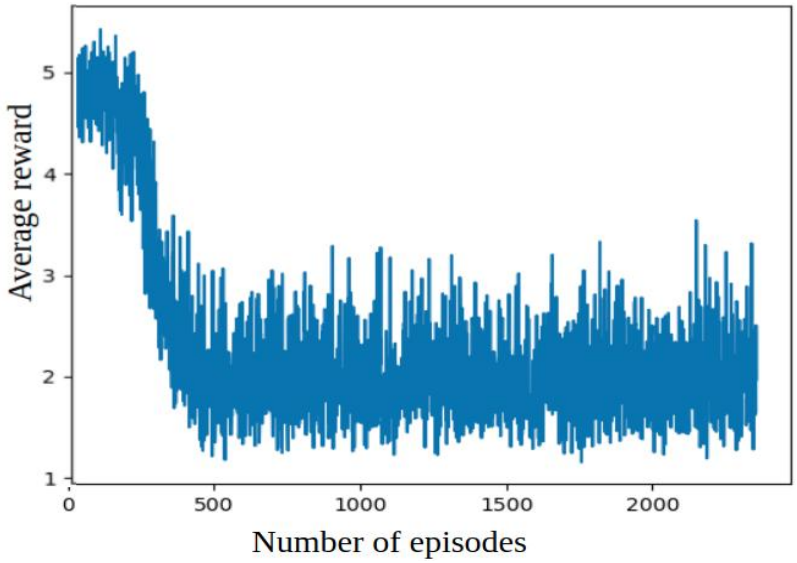


Figure 4.4: Dropout technique

Generally, the dropout technique is used for the regularization of neural network models. We applied the dropout of 0.3 (average of the trial values) for the input and hidden layers. For this technique, randomly selected neurons are ignored during training for better generalization and are less likely to overfit the training data. That means, their contribution to the activation of downstream neurons is removed temporarily on the forward pass, and any weight updates are not applied to the neuron on the backward pass. In our case, this technique does not influence positively, so we guess that the ignored neurons that contain valuable information or inaccurate tuning of the dropout for the training network.

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 $Q(S_t) \in R^4$.

4.3 Simulation results

First of all, we compared the performance of ten models which have the highest reward in different settings of the hyperparameter and network architechture, from which one of best performing model is selected, as shown in figure 4.5 and 4.6.

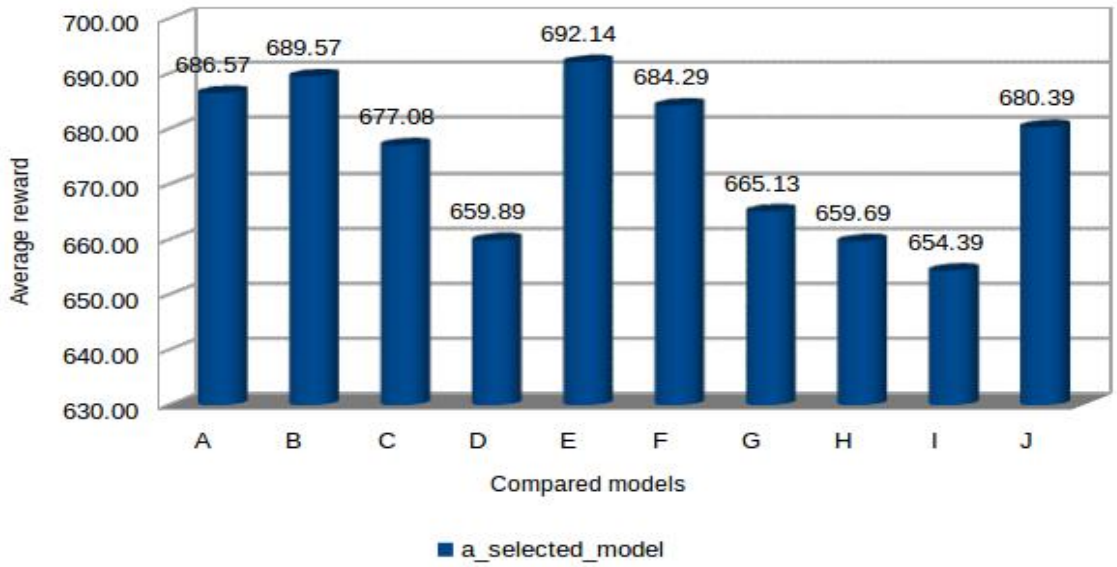


Figure 4.5: Comparison of the models with high reward

The reward average reward value of the compared models was 674.92. Moreover, model E have a most highest reward, 692.14. Therefore, before selecting certain model, we have to evaluate the performance of these models first.

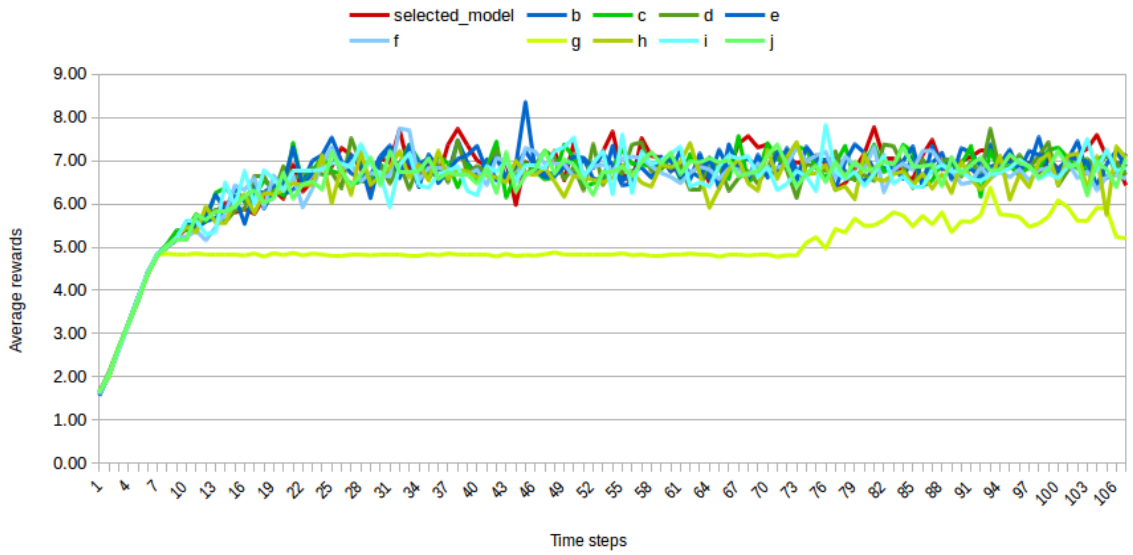


Figure 4.6: Comparison of the models with high reward

From the result of performance evaluation of compared models, it can be observed that the best model can't be directly chosen from the highest rewarded model. Because Model E provides the highest reward, as shown in figure 4.5, but the performance of that model (i.e., reward is 6.51) is worse than our selected model (i.e., reward is 6.58). In this figure, the x-axis holds the selected time steps and the y-axis holds the average reward for each compared model. It can be observed that the best model can not be directly chosen from the highest-rewarded model, as attached in appendix B. Where, we can see that model b provides the highest reward but the performance (i.e., average reward) of that model is worse than our selected model. Except for model g, other models achieve similar performance with each other. Performance of the obtained DDQN model in terms of average throughput, shown in figure 4.7, the horizontal axis is the number of episodes and the vertical axis is the average reward.

When the average reward (system throughput) converges, the agent has learned the assumed environment and is able to 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

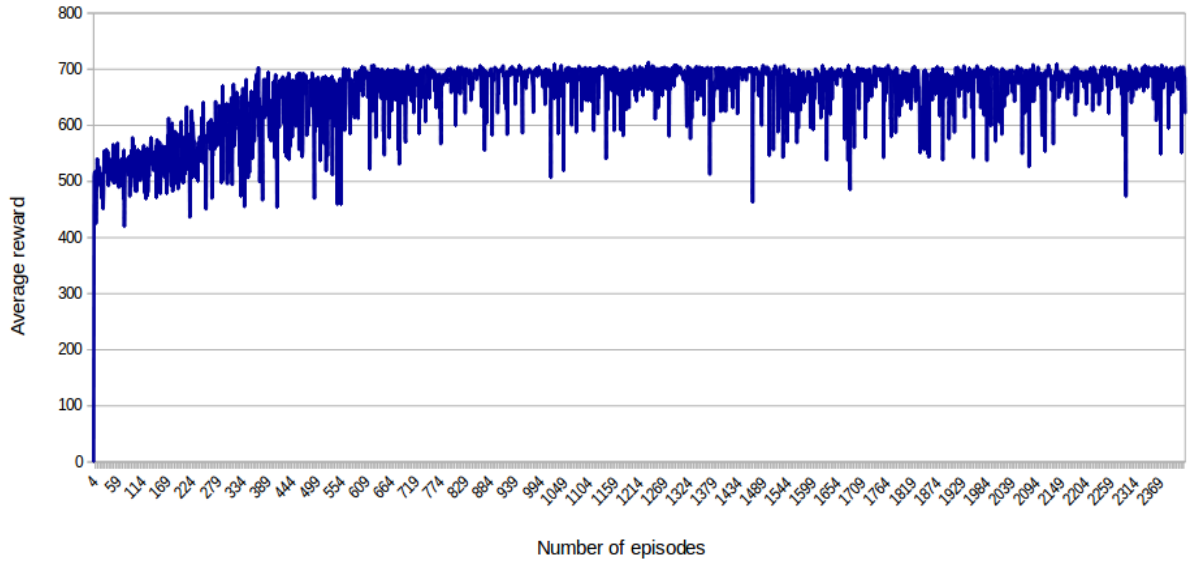


Figure 4.7: Performance of the obtained DDQN model in terms of average throughput

agent obtains small and random rewards. On the other hand, as the agent learns the environment and the value of ϵ decreases, the exploitation phase increases. 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 is able to receive the highest throughput from the available AP or eNBs in the assumed environment. The model reaches the optimal point at 350-400 episodes, where the agent gets to keep a stable reward. 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) is able to assign suitable channels for each AP/eNBs in the proposed environment.

Then the training stability of our obtained model is compared with the other eight selected models, as shown in figure 4.8.

These eight models are also trained on the same settings of hyperparameters and network architecture as the obtained model. When comparing these models, each training is run for the same number of episodes to collect the average performance for a fair comparison. In other words, for every model, the average reward is calculated

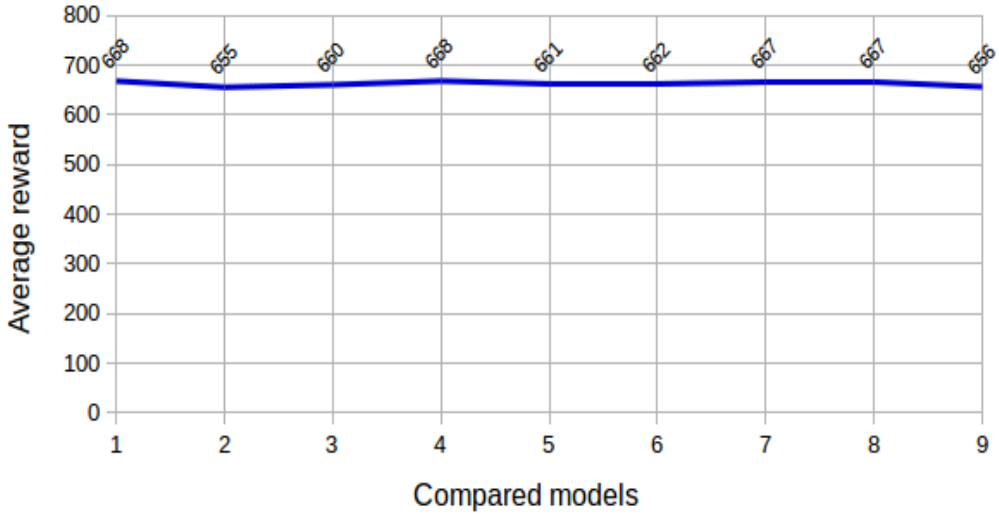


Figure 4.8: The average reward of compared models

for the same number of iterations (i.e., 2418 episodes). Note that the first one is our selected model on the horizontal axis and on 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) with respect to 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 sections 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 4.9.

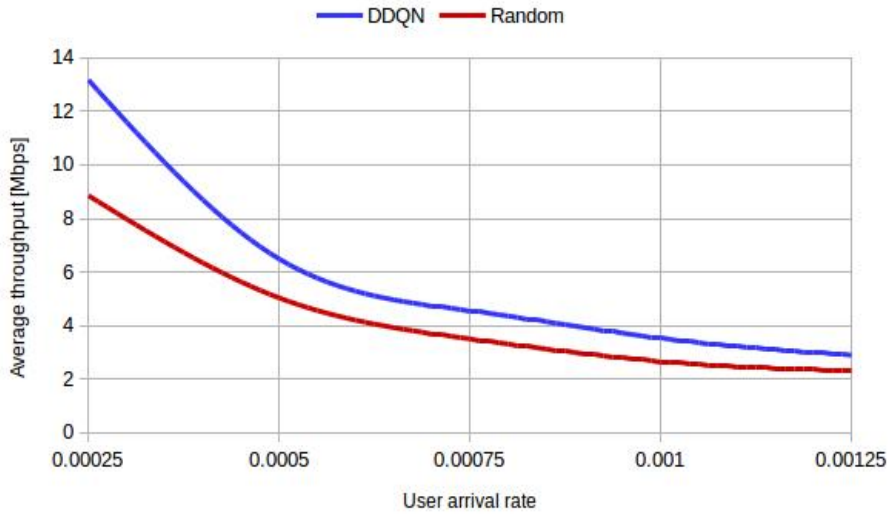


Figure 4.9: Comparison of average throughput in different arrival rates

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

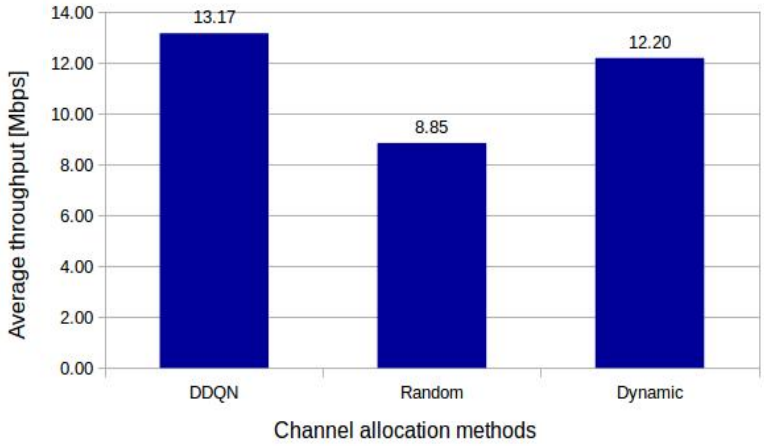


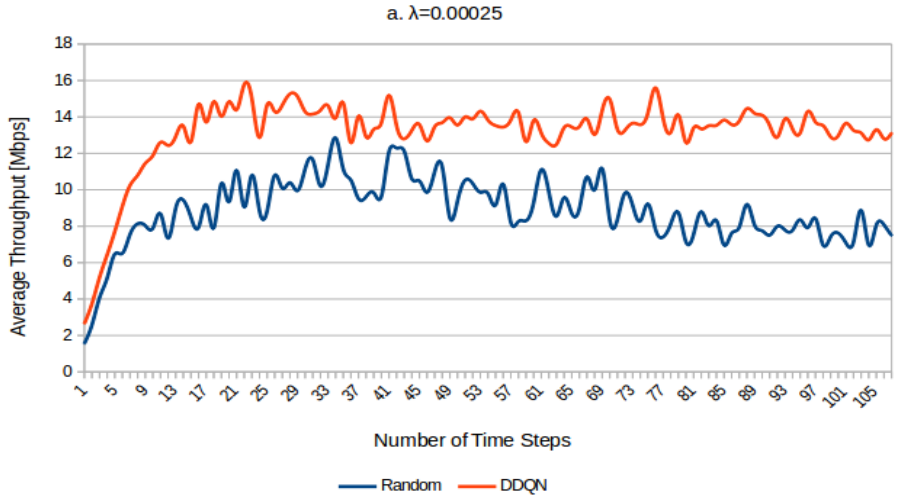
Figure 4.10: Comparison of the proposed method and the existing methods

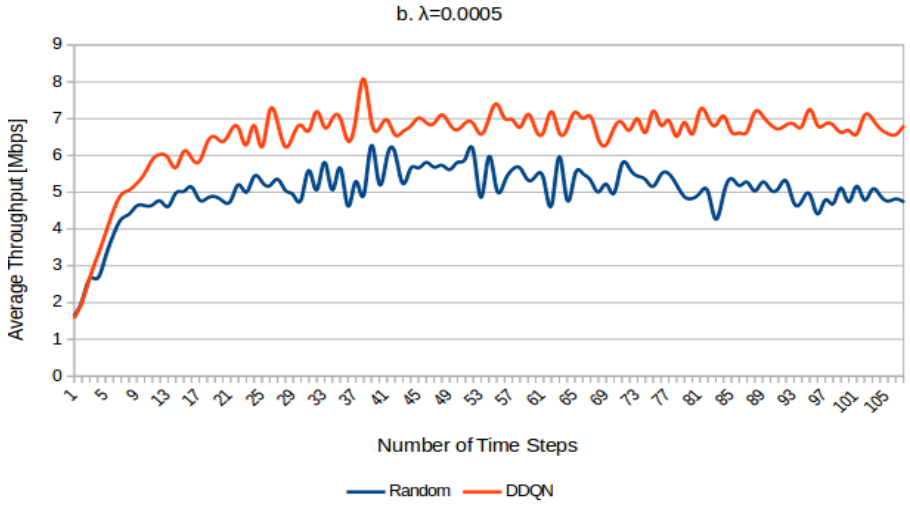
In addition, we compared the average throughput of our proposed method and other existing methods, (dynamic and random) when user arrival rate is 0.00025, as shown in figure 4.10. The dynamic channel assignment is executed at the same time interval of 300 sec as the assumed environment. Whereon the fewest used channel is dynamically assigned to a Wi-Fi AP based on the number of channels assigned

to APs and BSs whose coverage areas overlap. This method assumes that channels are assigned so as to avoid interference as much as possible based on the number of users. The numerical results show that our proposed DDQN algorithm provides the average throughput from 7.37% to 32.8% in 0.00025 of user arrival rate compared to the random and dynamic channel assignment methods. From the obtained 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.





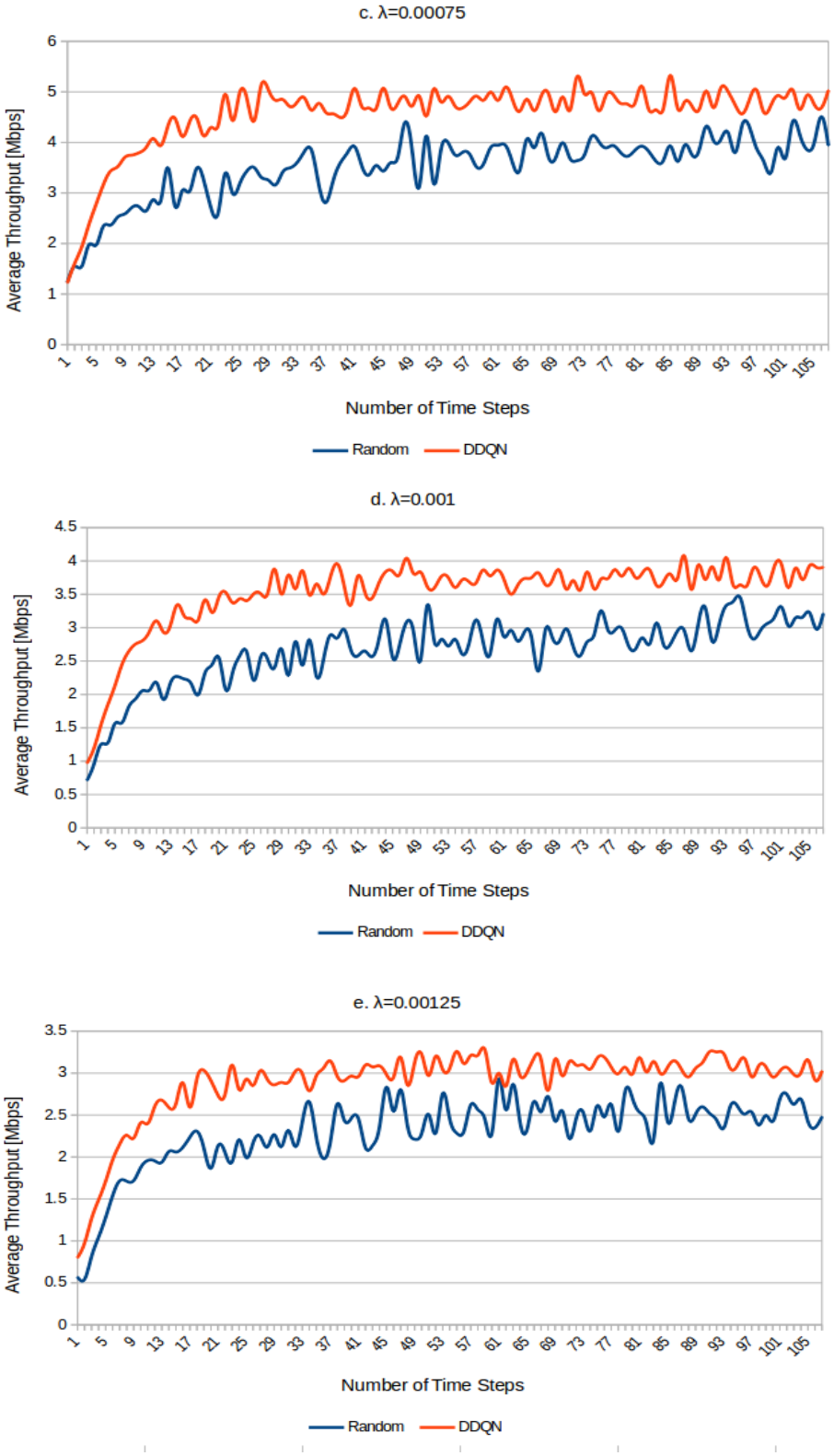


Figure 4.11: Validation results in different user arrival rates

After the training process, we evaluated the performance of the trained agent for 107 time steps, to confirm how well it has generalized to assign channels to the selected time steps for an episode in the proposed environment under the different user arrival rates, as represented in figure 4.11.

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.

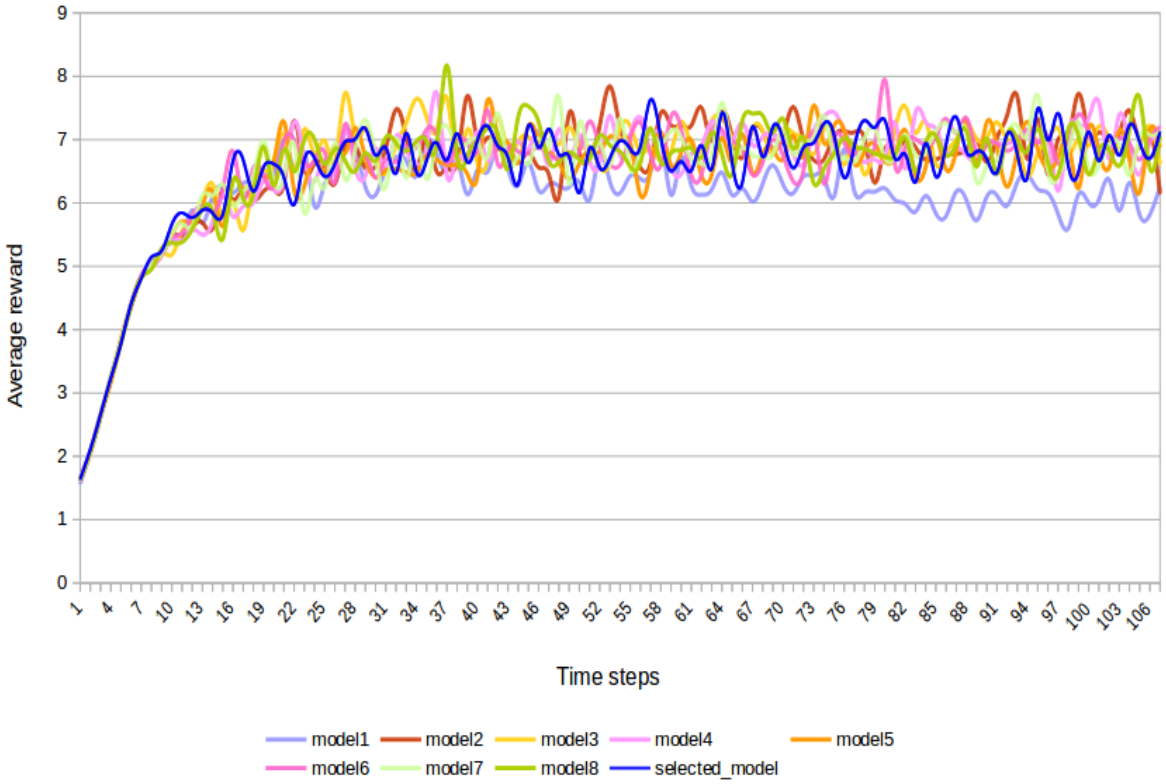


Figure 4.12: Comparison of stability for the obtained models

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 4.12.

In this figure, the horizontal axis is the number of time steps and the vertical axis is the average reward for different models. In terms of average reward, our agent obtains a minimum score of 1.69, a maximum score of 7.83, and the averaged reward of around 6.75. The averaged score of 6.75 is remarkably higher than 4.5 compared with the

average score of the random method, as shown in Figure 4.5(b). 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.

.

Chapter 5

Conclusion and Future Works

In this work, we proposed to improve the average throughput in densely deployed cellular/Wi-Fi heterogeneous wireless networks by Deep Reinforcement Learning based channel assignment method.

First of all, we have analyzed the different spectrum sharing techniques and the coexistence scenarios between cellular and Wi-Fi networks for further investigation.

Then, 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 based on LBT spectrum sharing mechanism.

Moreover, we have investigated the different algorithms of the Reinforcement Learning method for the RRAM problems. From these algorithms, we have developed our own DDQN algorithm for efficient channel allocation problem. After that we have trained DDQN agent based on the developed environment. Additionally, based on the developed environment, the training data was generated which also can be used for training DRL based models by an offline manner.

Regarding the trained 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. Finally, we evaluated the generalization performance and the stability 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. Moreover, we intend to investigate a distributed technique where each user can learn a policy about channel selection independently. It means implementing a DQN algorithm for each user (i.e., multi-agent) independently. Then, users are able to learn their channel selection policies simultaneously and prevent interference based on the knowledge accumulated from observations and rewards.

.

.

Acknowledgements

First of all, I would like to thank my supervisor, Professor Kazuhiko Kinoshita the Department of Information Science and Intelligent Systems at the Graduate School of Advanced Technology and Science, Tokushima University for all his guidance, continuous support, and invaluable advice that he has provided me during my Ph.D. degree. I could not have completed my research study and dissertation without his guidance and advice.

I would like to express a special thank you to Associate Professor Kenji Ikeda of the Department of Information Science and Intelligent Systems in the Graduate School of Advanced Technology and Science, Tokushima University for his support during the study.

I would like to give a special thanks of gratitude to my Professors, Otgonbayar Bataa and Khishigjargal Gonchigsumlaa of the School of Information and Communication Technology, Mongolian University of Science and Technology for their support and encouragement. They gave me the wonderful opportunity to study as a Doctoral student in Japan and find a supervisor.

I would like to express my sincere appreciation to the Higher Engineering Education Development Project, named Mongolia Japan Engineering Education Development (MJEED). I received a great opportunity to study doctoral program in Japan and full financial support from the project.

Words cannot express my gratitude to my mother, Dariimaa, have been providing me with love and endless support for the whole of my life. Most importantly, I would like to extend my sincere thanks to my husband, Bayarbaatar, and my two lovely children, Khanbat and Badamzul have been offering me a lot of love and unending inspiration. I cannot imagine myself finishing my Ph.D. without them. I would also like to thank my siblings (Baigalmaa, Enkhbayar, Purevkhuu, and Dolgormaa) and their families for their help and support.

I would like to extend my sincere thanks to Uuganbayar Davaasuren and Gantulga who generously provided knowledge and expertise. I am also thankful to fellow students at the laboratory for their assistance whenever.

I would like to express my deepest appreciation to Japan Educational Exchanges and Services for supporting me the JEES-MUFG scholarship.

Finally, I would like to thank all the teachers and instructors who taught me in elementary school, high school, and university for their help and motivation.

DEDICATION

I dedicate this dissertation to
my children, KHANBAT and BADAMZUL.

.

Bibliography

[1] Imt traffic estimates for the years 2020 to 2030. *Report. ITU-R M.2370-0*, pages 16–24, 7/2015.

[2] Signals Research Group. The prospect lte wi-fi sharing unlicensed spectrum. *White Paper*, 2015.

[3] Mamta Agiwal, Abhishek Roy, and Navrati Saxena. Next generation 5g wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, pages 1617–1655, 2016.

[4] Feng Hu, Bing Chen, and Kun Zhu. Full spectrum sharing in cognitive radio networks toward 5g: A survey. *IEEE Access*, 6, 2018.

[5] Yubing Jian, Chao-Fang Shih, Bhuvana Krishnaswamy, and Raghupathy Sivakumar. Coexistence of wi-fi and laa-lte: Experimental evaluation, analysis and insights. In *2015 IEEE international conference on communication workshop (ICCW)*.

[6] Gaurang Naik, Jinshan Liu, and Jung-Min Jerry Park. Coexistence of wireless technologies in the 5 ghz bands: A survey of existing solutions and a roadmap for future research. *IEEE communications surveys & tutorials*, 20(3):1777–1798, 2018.

[7] Shaoyi Xu, Yan Li, Yuan Gao, Yang Liu, and Haris Gačanin. Opportunistic coexistence of lte and wifi for future 5g system: Experimental performance evaluation and analysis. *Ieee Access*, 6:8725–8741, 2017.

[8] Mohammed Hirzallah, Marwan Krunz, Balkan Kecicioglu, and Belal Hamzeh. 5g new radio unlicensed: Challenges and evaluation. *IEEE Transactions on Cognitive Communications and Networking*, 7(3):689–701, 2020.

[9] Bolin Chen, Jiming Chen, Yuan Gao, and Jie Zhang. Coexistence of lte-laa and wi-fi on 5 ghz with corresponding deployment scenarios: A survey. *IEEE Communications Surveys & Tutorials*, 19(1):7–32, 2016.

[10] Andra M Voicu, Ljiljana Simić, and Marina Petrova. Survey of spectrum sharing for inter-technology coexistence. *IEEE Communications Surveys & Tutorials*, 21(2):1112–1144, 2018.

[11] Le Liang, Hao Ye, Guanding Yu, and Geoffrey Ye Li. Deep-learning-based wireless resource allocation with application to vehicular networks. *IEEE*, 108(2):341–356, 2019.

[12] Abdulmalik Alwarafy, Mohamed Abdallah, Bekir Sait Ciftler, Ala Al-Fuqaha, and Mounir Hamdi. Deep reinforcement learning for radio resource allocation and management in next generation heterogeneous wireless networks: A survey. *TechRxiv*, 05 2021.

[13] Kota Nakashima, Shotaro Kamiya, Kazuki Ohtsu, Koji Yamamoto, Takayuki Nishio, and Masahiro Morikura. Deep reinforcement learning-based channel allocation for wireless lans with graph convolutional networks. *IEEE Access*, 8:31823–31834, 2020.

[14] C Capretti, Francesco Gringoli, Nicolò Facchi, and Paul Patras. Lte/wi-fi coexistence under scrutiny: An empirical study. In *10 ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation, and Characterization*, pages 33–40, 2016.

[15] Sabin Bhandari and Sangman 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, 2019.

[16] Erika Almeida, André M Cavalcante, Rafael CD Paiva, Fabiano S Chaves, Fuad M Abinader, Robson D Vieira, Sayantan Choudhury, Esa Tuomaala, and

Klaus Doppler. Enabling lte/wifi coexistence by lte blank subframe allocation. In *2013 IEEE International Conference on Communications (ICC)*, pages 5083–5088, 2013.

[17] Intel Corporation Sasha Sirotkin. Lte-wlan aggregation (lwa): Benefits and deployment considerations. *White Paper*, 2016.

[18] Gaurang Naik, Jung-Min Park, Jonathan Ashdown, and William Lehr. Next generation wi-fi and 5g nr-u in the 6 ghz bands: Opportunities and challenges. *IEEE Access*, 8:153027–153056, 2020.

[19] Vasilis Maglogiannis, Dries Naudts, Adnan Shahid, and Ingrid Moerman. An adaptive lte listen-before-talk scheme towards a fair coexistence with wi-fi in unlicensed spectrum. *Telecommunication Systems*, 68(4):701–721, 2018.

[20] Kazuhiko Kinoshita, Kazuki Ginnan, Keita Kawano, Hiroki Nakayama, Tsunemasa Hayashi, and Takashi Watanabe. Channel assignment and access system selection in heterogeneous wireless network with unlicensed bands. In *2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pages 96–101, 2020.

[21] Jacek Wszołek, Szymon Ludyga, Wojciech Anzel, and Szymon Szott. Revisiting lte laa: Channel access, qos, and coexistence with wifi. *IEEE Communications Magazine*, 59(2):91–97, 2021.

[22] Cheng Chen, Rapeepat Ratasuk, and Amitava Ghosh. Downlink performance analysis of lte and wifi coexistence in unlicensed bands with a simple listen-before-talk scheme. In *IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2015.

[23] Yemeserach Mekonnen, Muhammad Haque, Imtiaz Parvez, Amir Moghadasi, and Arif Sarwat. Lte and wi- fi coexistence in unlicensed spectrum with application to smart grid: a review. In *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pages 1–5, 2018.

[24] Sandra Lagen, Lorenza Giupponi, Sanjay Goyal, Natale Patriciello, Biljana Bojović, Alpaslan Demir, and Mihaela Beluri. New radio beam-based access to

unlicensed spectrum: Design challenges and solutions. *IEEE Communications Surveys & Tutorials*, 22(1):8–37, 2019.

[25] Moawiah Alhulayil and Miguel López-Benítez. Novel laa waiting and transmission time configuration methods for improved lte-laa/wi-fi coexistence over unlicensed bands. *IEEE Access*, 8:162373–162393, 2020.

[26] Yuan Gao, Xiaoli Chu, and Jie Zhang. Performance analysis of laa and wifi coexistence in unlicensed spectrum based on markov chain. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2016.

[27] Zhenzhou Tang, Xuesheng Zhou, Qian Hu, and Guanding Yu. Throughput analysis of laa and wi-fi coexistence network with asynchronous channel access. *IEEE Access*, 6:9218–9226, 2018.

[28] Rony Kumer Saha. 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*, 2021.

[29] Sandra Lagen, Natale Patriciello, and Lorenza Giupponi. Cellular and wi-fi in unlicensed spectrum: Competition leading to convergence. In *2020 2nd 6G Wireless Summit (6G SUMMIT)*, pages 1–5, 2020.

[30] Ursula Challita and David Sandberg. Deep reinforcement learning for dynamic spectrum sharing of lte and nr. In *ICC 2021-IEEE International Conference on Communications*, pages 1–6. IEEE, 2021.

[31] Vasilis Maglogiannis, Dries Naudts, Adnan Shahid, and Ingrid Moerman. A q-learning scheme for fair coexistence between lte and wi-fi in unlicensed spectrum. *IEEE Access*, 6:27278–27293, 2018.

[32] Adam Dziedzic, Vanlin Sathya, Muhammad Iqbal Rochman, Monisha Ghosh, and Sanjay Krishnan. Machine learning enabled spectrum sharing in dense lte-u/wi-fi coexistence scenarios. *IEEE Open Journal of Vehicular Technology*, 1:173–189, 2020.

[33] Vasilis Maglogiannis, Adnan Shahid, Dries Naudts, Eli De Poorter, and Ingrid Moerman. Enhancing the coexistence of lte and wi-fi in unlicensed spectrum through convolutional neural networks. *IEEE Access*, 7:28464–28477, 2019.

[34] Shangxing Wang, Hanpeng Liu, Pedro Henrique Gomes, and Bhaskar Krishnamachari. Deep reinforcement learning for dynamic multichannel access in wireless networks. *IEEE Transactions on Cognitive Communications and Networking*, 4(2):257–265, 2018.

[35] Ursula Challita, Li Dong, and Walid Saad. Proactive resource management for lte in unlicensed spectrum: A deep learning perspective. *IEEE Transactions on Wireless Communications*, 17(7):4674–4689, 2018.

[36] Oshri Naparstek and Kobi Cohen. Deep multi-user reinforcement learning for distributed dynamic spectrum access. *IEEE transactions on wireless communications*, 18(1):310–323, 2018.

[37] Shaoyang Wang and Tiejun Lv. Deep reinforcement learning based dynamic multichannel access in hetnets. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2019.

[38] Haixia Peng and Xuemin Shen. Deep reinforcement learning based resource management for multi-access edge computing in vehicular networks. *IEEE Transactions on Network Science and Engineering*, 7(4):2416–2428, 2020.

[39] Najem N Sirhan and Manel Martinez Ramon. Cognitive radio resource scheduling using multi agent qlearning for lte. *arXiv preprint arXiv:2205.02765*, 2022.

Publications

Main

- Bayarmaa Ragchaa, and Kazuhiko Kinoshita, "Spectrum Sharing between Cellular and Wi-Fi Networks Based on Deep Reinforcement Learning," International Journal of Computer Networks & Communications (IJCNC) Vol.15, No.1, January 2023.

Others

- Bayarmaa Ragchaa, and Kazuhiko Kinoshita, "Deep Reinforcement Learning Based Channel Assignment for Cellular and Wi-Fi Heterogeneous Network in Unlicensed Bands," APNOMS 2022 (The 23 rd Asia-Pacific Network Operations and Management Symposium), Sep 28-30, 2022.
- Bayarmaa Ragchaa, Kazuhiko Kinoshita, Bataa Otgonbayar and Erdenebayar "Optimal channel assignment in LTE/WiFi heterogeneous network," to eNATION-ICT100, June 17, 2021.
- Bayarmaa Ragchaa, and Kazuhiko Kinoshita, "An Efficient Channel Assignment based on Deep Reinforcement Learning in Heterogeneous Wireless Network with Unlicensed Bands," RISING, Nov 15, 2021.
- Bayarmaa Ragchaa, and Kazuhiko Kinoshita, "A Performance Evaluation on Channel Assignment based on Deep Reinforcement Learning in Heterogeneous Wireless Network with Unlicensed Bands," RISING, Oct 31, 2022.

.

APPENDIXES

Appendix A: Acronyms

3gpp 3rd Generation Partnership Project

5G Fifth-Generation

A3C Asynchronous Advantage Actor Critic

ABS Almost Blank Subframe

ACK Acknowledgment

AI Artificial Intelligence

AIFS Arbitration Inter-Frame Space

APs Access Points

ARQ Automatic Repeat Request

BS Base Station

CA Carrier Aggregation

CBRS Citizens Broadband Radio Service

CCA Clear Channel Assessment

CN Core Network

CNN Conventional Neural Network

COT Channel Occupancy Time

CR Cognitive Radio

CS Carrier Sense

CSAT Carrier Sensing Adaptive Transmission

CSMA/CS Carrier Sensing Multiple Access with Collision Avoidance

CW Contention Window

DC Dual Connectivity

DCA Dynamic Channel Access

DCF Distributed Coordination Function

DDPG Deep Deterministic Policy Gradient

DDQN Double Deep Q Networks

DFS Dynamic Frequency Selection

DIFS DCF Interframe Space

DL Deep Learning

DNN Deep Neural Network

DQN Deep Q Network

DRL Deep Reinforcement Learning

DRS Discovery Reference Signal

DSRC Dedicated Short-Range Communications

ECCA Extended Clear Channel Assessment

ED Energy Detection

EIRP Effective Isotropic Radiated Power

eLAA enhanced LAA

eNBs eNodeBs

EPC Evolved Packet Core

ETSI European Telecommunications Standards Institute

FBE Frame Based Equipment

FCC Federal Communications Commission

feLAA further enhanced LAA

HARQ Hybrid Automatic Repeat Request

IoT Internet of Things

ISM Industrial Scientific Medical

LAA License Assisted Access

LBA Load Based Equipment

LBT Listen Before Talk

LOS Line Of Sight

LTE Long Term Evolution

LWA LTE Wi-Fi Aggregation

MAC Medium Access Control

MBSFN Multimedia Broadcast Single Frequency Network

MCOT Maximum Channel Occupancy Time

MCA Multi Channel Access

MCTS Monte Carlo Tree Search

MEC Multi-access Edge Computing

MeNB Macro eNodeB

MIMO Multiple Input Multiple Output

MDP Markov Decision Process

ML Machine Learning

NN Neural Network

NR-U New Radio Unlicensed

OFDMA Orthogonal Frequency Division Multiple Access

PCell Primary Cell

PDCH Physical Broadcast Channel

PHY Physical Layer

PSS Primary Synchronization Signal

QoS Quality of Service

RAN Radio Access Network

RAT Radio Access Technology

RLU Rectifier Linear Units

RL Reinforcement Learning

RRAM Radio Resource Allocation and Management

SCell Secondary Cell

SDL Supplementary Downlink

SBSs Small Base Stations

SSB Synchronization Signal Block

SSS Secondary Synchronization Signal

TCP Transmission Control Protocol

TD Temporal Difference

TDD Time Division Duplex

TXOP Transmission Opportunity

U-NII Unlicensed National Information Infrastructure

UE User Equipment

WLAN Wireless Local Area Network

Appendix B: Comparison of model's reward

	Models reward (λ=0.0005)									
	selectedmodel	b	c	d	e	f	g	h	I	j
TS	706.95	712.45	704.11	706.04	705.69	704.78	704.72	708.63	705.21	705.42
1	**1.68**	1.60	1.64	1.65	1.57	1.61	1.66	1.65	1.64	1.63
2	**2.02**	2.06	2.01	2.12	2.06	2.04	2.02	2.08	2.06	2.06
3	**2.63**	2.66	2.66	2.69	2.66	2.65	2.66	2.69	2.63	2.69
4	**3.24**	3.23	3.25	3.21	3.24	3.26	3.23	3.20	3.21	3.24
5	**3.83**	3.84	3.81	3.80	3.76	3.82	3.81	3.77	3.82	3.76
6	**4.37**	4.39	4.41	4.42	4.42	4.43	4.38	4.38	4.41	4.38
7	**4.85**	4.82	4.82	4.82	4.85	4.81	4.83	4.84	4.86	4.80
8	**5.00**	5.02	5.07	5.00	5.00	5.03	4.85	5.00	5.01	5.04
9	**5.25**	5.18	5.39	5.26	5.18	5.22	4.82	5.23	5.22	5.19
10	**5.42**	5.59	5.39	5.19	5.58	5.24	4.83	5.48	5.60	5.18
11	**5.36**	5.43	5.76	5.66	5.54	5.38	4.86	5.34	5.60	5.74
12	**5.84**	5.60	5.59	5.69	5.69	5.16	4.84	5.94	5.30	5.63
13	**5.80**	5.67	6.25	5.87	6.22	5.46	4.82	5.57	5.35	5.82
14	**6.03**	5.87	6.37	5.58	5.85	5.77	4.83	5.55	6.49	5.83
15	**5.80**	6.34	6.15	5.85	6.17	6.43	4.83	6.07	5.96	5.93
16	**5.89**	6.05	6.11	5.88	5.54	6.32	4.81	6.22	6.76	6.18
17	**5.77**	6.45	6.18	6.63	6.28	6.60	4.86	5.79	6.00	6.31
18	**6.13**	5.90	6.54	6.66	6.28	5.93	4.78	6.25	6.78	6.03
19	**6.32**	6.30	6.35	6.22	6.42	6.65	4.86	6.22	6.57	6.13
20	**6.12**	6.21	6.19	6.87	6.45	6.31	4.82	6.32	6.31	6.72
21	**6.91**	6.33	7.41	6.58	7.30	6.65	4.87	6.48	6.76	6.12
22	**6.28**	6.50	6.45	6.69	6.42	5.92	4.81	6.43	6.77	6.46
23	**6.53**	6.99	6.76	6.83	6.71	6.35	4.85	6.45	6.77	6.52
24	**6.85**	7.13	7.08	6.67	6.94	6.94	4.83	7.12	6.79	6.33
25	**6.85**	7.53	6.75	6.73	7.28	7.30	4.81	6.02	6.94	7.20
80	**7.17**	6.52	6.90	6.67	7.18	6.81	5.49	7.16	6.41	6.76
81	**7.77**	6.84	7.37	6.53	6.79	7.32	5.51	6.59	6.70	6.81
82	**7.06**	6.84	6.74	7.37	7.02	6.27	5.63	6.53	7.13	6.98
83	**7.04**	6.84	6.87	7.33	6.80	6.79	5.81	6.68	6.47	6.91
84	**6.66**	6.58	7.36	6.85	7.00	6.94	5.73	6.75	6.64	7.32
85	**6.57**	7.12	7.20	6.37	7.33	6.78	5.48	6.43	6.39	6.87
86	**7.13**	6.50	6.40	7.03	6.41	7.22	5.73	6.70	6.40	6.86
87	**7.48**	6.94	6.54	6.73	7.34	7.23	5.53	6.35	6.79	6.92
88	**6.86**	7.18	6.97	7.01	6.80	6.91	5.80	6.69	6.90	6.40
89	**6.70**	6.56	6.90	7.04	7.00	6.78	5.35	6.25	6.74	7.08
90	**6.99**	7.29	6.89	6.63	6.73	6.46	5.60	6.98	6.77	6.68
91	**7.09**	7.19	7.34	6.98	6.81	6.50	5.58	6.59	6.58	6.90
92	**7.24**	6.86	6.17	6.70	6.61	6.85	5.74	6.32	6.53	7.07
93	**7.23**	7.38	7.23	7.73	6.86	6.59	6.37	6.58	6.68	6.76
94	**6.81**	6.91	6.72	6.77	6.93	6.75	5.76	7.11	6.73	6.79
95	**7.07**	6.99	6.89	6.82	7.25	6.62	5.74	6.10	6.83	7.14
96	**6.92**	6.80	6.66	7.14	6.75	6.80	5.70	6.68	7.02	6.94
97	**6.95**	6.85	6.86	6.81	7.22	6.54	5.47	6.39	6.86	6.88
98	**6.65**	7.54	6.77	7.11	7.13	6.79	5.55	6.98	6.57	6.62
99	**7.04**	6.73	7.25	7.42	7.08	6.83	5.71	7.16	6.70	6.81
100	**6.81**	6.82	7.30	6.43	6.72	6.55	6.08	6.43	6.59	7.22
101	**6.97**	7.02	7.06	6.77	7.09	6.97	5.92	7.11	6.86	6.86
102	**7.15**	7.45	6.81	6.96	7.03	6.59	5.61	7.16	6.92	6.91
103	**7.27**	6.70	6.85	7.04	6.97	6.75	5.60	6.35	7.50	6.20
104	**7.59**	6.87	7.01	6.84	6.43	6.31	5.91	7.10	6.77	6.97
105	**7.09**	6.60	7.15	7.13	6.92	6.63	5.89	5.75	7.25	6.80
106	**6.96**	6.78	6.75	6.96	7.22	6.66	5.23	7.33	6.76	6.39
107	**6.43**	7.16	6.80	6.67	6.86	6.58	5.21	7.06	6.84	7.05
Avg	**6.58**	6.51	6.52	6.48	6.55	6.46	4.97	6.35	6.46	6.45

Appendix C: Input data for Wi-Fi

Users	AP ID	Max Capacity	User Throughput	Assigned channels	User's Location ID	AP's Loc ID
0	11	40	9.00	3	66	47
1	92	40	5.20	3	256	84
2	63	40	6.75	3	123	15
3	86	40	13.50	1	185	85
4	54	40	13.50	1	261	123
5	63	40	13.50	3	107	15
6	63	40	13.50	3	35	15
7	16	40	6.50	1	227	111
8	63	40	13.50	3	17	15
9	19	40	13.00	2	44	20
10	79	40	6.50	3	147	28
11	72	40	9.00	0	259	110
12	38	40	13.50	3	17	52
13	86	40	13.50	1	187	85
14	19	40	13.00	2	44	20
15	54	40	13.50	1	242	123
16	19	40	13.00	2	64	20
17	97	40	9.00	0	179	12
18	63	40	13.50	3	106	15
19	86	40	6.75	1	259	85
20	19	40	6.50	2	5	20
21	70	40	8.67	1	39	62
22	54	40	13.50	1	243	123
23	86	40	13.50	1	187	85
24	64	40	9.00	2	2	49
25	92	40	5.20	3	221	84
26	19	40	5.20	2	25	20
27	86	40	6.75	1	167	85
28	72	40	9.00	0	259	110
29	56	40	4.33	0	166	87
30	64	40	9.00	2	56	49
31	86	40	13.50	1	167	85
32	58	40	13.50	1	39	94
33	34	40	9.00	3	219	126
34	72	40	9.00	0	260	110
35	49	40	13.50	0	144	24
36	70	40	13.00	1	0	62
37	62	40	13.50	0	67	88
38	74	40	9.00	3	212	1
39	86	40	13.50	1	148	85
40	66	40	13.50	0	156	79
41	63	40	13.50	3	107	15
42	53	40	13.50	1	150	120
43	75	40	9.00	1	172	134
44	90	40	13.50	0	123	50
45	66	40	13.50	0	45	79
46	72	40	9.00	0	259	110
47	51	40	9.00	0	150	22
48	66	40	13.50	0	32	79
184	34	40	9.00	3	183	126

Appendix D: Input data for LTE

Users	BS's ID	Max Capacity	User Throughput	Assigned channels	User's Location 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	2	188	104
7	10004	75	9.4	1	212	101
8	10000	75	11	0	91	26
9	10001	75	7.83	1	63	29
10	10000	75	33	0	58	26
11	10002	75	6.6	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	0	130	26
17	10000	75	6.6	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	9.4	1	230	101
23	10006	75	6.71	0	226	107
24	10001	75	9.4	1	64	29
25	10000	75	11	0	54	26
26	10005	75	9	2	205	104
27	10000	75	16.5	0	110	26
28	10000	75	33	0	92	26
29	10003	75	9.4	1	128	93
30	10006	75	7.83	0	227	107
31	10004	75	7.83	1	212	101
32	10006	75	11.75	0	226	107
33	10002	75	6.6	2	140	32
34	10001	75	11.75	1	96	29
35	10004	75	9.4	1	159	101
36	10004	75	11.75	1	268	101
37	10004	75	11.75	1	160	101
38	10006	75	15.67	0	264	107
39	10006	75	23.5	0	265	107
40	10003	75	7.83	1	234	93
41	10000	75	8.25	0	91	26
42	10004	75	11.75	1	179	101
43	10003	75	9.4	1	162	93
44	10002	75	5.5	2	84	32
45	10004	75	15.67	1	196	101
46	10003	75	11.75	1	234	93
47	10000	75	8.25	0	128	26
48	10002	75	6.6	2	122	32
464	10000	75	33	0	5	26