

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Deep Q-Learning-Based Dynamic Spectrum Sharing Between Licensed 5G New Radio and Wi-Fi-Inspired Cognitive Radio Secondary Users

A Thesis Submitted to School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Telecom Information System

By

Melikte Yohannes

Advisor

Yihenew Wondie (Ph.D)

May 7, 2026

Addis Ababa, Ethiopia



Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Deep Q-Learning-Based Dynamic Spectrum Sharing Between Licensed 5G New Radio and Wi-Fi-Inspired Cognitive Radio Secondary Users

By: Melikte Yohannes

APPROVED BY BOARD OF EXAMINERS

Name	Date:	Signature:
(Dean, School of Graduate Committee)		
(Advisor)		
(Internal Examiner)		
(External Examiner)		

Declaration

I, the undersigned, declare that this MSc thesis entitled **" Deep Q-Learning-Based Dynamic Spectrum Sharing Between Licensed 5G New Radio and Wi-Fi-Inspired Cognitive Radio Secondary Users"** is the result of my independent research work and has not been presented for fulfillment of any degree in this or any other university, and all sources and materials used for this thesis are properly acknowledged.

Name of the student	Date	Signature

This thesis has been submitted for examination with my approval as a university advisor.

Name of the Advisor	Date	Signature

Acknowledgement

I would like to express my sincere gratitude to my advisor, Dr.Yihenew Wondie, for his guidance, support, and constructive feedback throughout the development of this thesis. I also extend my appreciation to the School of Electrical and Computer Engineering at College of Technology and Built Environment for providing the academic environment necessary to complete this work. I am grateful to my examiners for their encouragement and patience. Their support has been really helpful.

Abstract

The increasing demand for wireless networks has made effective spectrum utilization imperative. One of the most recent technologies develped to boost capacity in order to satisfy the growing users demand is 5th generation network. Bandwidth is a crucial component of 5G's overall network performance. Network capacity can be increased by boosting spectrum efficiency allowing more data to be sent over the same bandwidth. In licensed bands, spectrum gaps are created because primary users do not utilize spectrum at all times due to the dynamic nature of 5th generation network. One important technique that allocates spectrum according to the network conditions is Dynamic Spectrum sharing enabling secondary users to access these idle licensed spectrum while keeping the priority of the licensed users. primary users share their channels with secondary users in ,cognitive radio, a crucial technology that uses dynamic spectrum allocation to manage spectrum scarcity in a network.In this paper a slot level simulation is performed between 5G New Radio operating at n40 bandwith bandwidth of 20MHZ and Deep Q-Learning controlled cognitive radio based WI-FI devices that access idle spectrum while keeping the primary users protected. Energy detection method with multi threshold sensing is applied. The primary user traffic is generated using a 4 State markov's chain with state update every 10ms. The secondary users have 1 access poInt and 10 secondary users and access channel opportunistically using 802.11ax carrier sensing. The simulation runs in 1 ms slot units, the secondary user's access point senses the primary user's channel and the deep Q-Learning agent decides the secondary user's action. The propagation environment is modeled using 3GPP based outdoor to indoor, and indoor office channel model. Indoor Line of sight /Non-Line of sight pathloss model is used for acess point-Secondary User and Secondary User-Secondary User.

Keywords: *5G Networks, Spectrum Management, Cognitive Radio, Energy Detection, Spectrum Sensing, Deep Q-Learning, Primary Users, Secondary Users*

Contents

Declaration ii

Acknowledgement iii

Abstract iv

List of Tables viii

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Problem Statement . . . 2
- 1.3 Research Objectives . . . 2
- 1.4 Research Questions . . . 2
- 1.5 Contribution . . . 3
- 1.6 Motivation of the Study and Scope . . . 3
- 1.7 Thesis Organization . . . 3

2 Literature Review 4
- 2.1 Introduction . . . 4
  - 2.1.1 Cognitive Radio Networks for Opportunistic Spectrum Access . . . 4
  - 2.1.2 Deep Reinforcement Learning Approaches and Resource Management 5
  - 2.1.3 5G NR and WI-FI Coexistence . . . 6
  - 2.1.4 Propagation Modeling . . . 7

3 System Modeling 9
- 3.1 Introduction . . . 9
- 3.2 Methodology . . . 9
- 3.3 System Modeling . . . 11
  - 3.3.1 Energy Detection Theory . . . 11
  - 3.3.2 The Detection-False Alarm Trade-off . . . 12
  - 3.3.3 Reinforcement Learning Fundamentals . . . 12
  - 3.3.4 Markov Chain Traffic Modeling . . . 12
  - 3.3.5 Deep Q-Learning Formulation . . . 14
  - 3.3.6 State Space . . . 14
  - 3.3.7 Action Space . . . 15
  - 3.3.8 Q-Value Function . . . 15
  - 3.3.9 Deep Q-Networks . . . 15
  - 3.3.10 DQL Target Value . . . 16
  - 3.3.11 Loss Function . . . 16
  - 3.3.12 LBT Protocol . . . 17

4 DQL-based spectrum sharing simulation Framework **18**
4.1 Simulation Parameters . . . 18
4.1.1 Primary User Traffic Model . . . 21
4.1.2 Secondary User Network Model . . . 21
4.1.3 Path-Loss and Propagation Model . . . 22
4.1.4 System Component Interactions . . . 23
4.1.5 Spectrum Sensing Configuration . . . 24

5 Results and Discussion **25**
5.1 Introduction . . . 25
5.1.1 Simulation Results . . . 25
5.1.2 Secondary User Throughput . . . 27
5.1.3 Queue Length and Delay Dynamics . . . 28
5.1.4 Primary User Protection . . . 29
5.1.5 Spectrum Utilization and Idle Sensing Fraction . . . 32

6 Conclusion and Future Work **40**
6.1 Conclusion . . . 40
6.2 Limitations . . . 40
6.2.1 Future Work . . . 41

Bibliography **42**

A Important MATLAB Codes **45**
A.1 Simulation Parameter Definition . . . 45
A.2 Primary User Markov Traffic Model . . . 46
A.3 Deep Q-Learning Configuration . . . 46
A.4 Reward Function . . . 47
A.5 Spectrum Sensing and Multi-Threshold Classification . . . 47
A.6 DQL State Formation . . . 48
A.7 Action Decoding . . . 48
A.8 Training Loop . . . 49
A.9 Result Figure Generation . . . 49

List of Figures

3.1 Flow Chart 10
3.2 System Model 11

5.1 DQL Reward Convergence Curve 26
5.2 Secondary User Throughput Curve 27
5.3 Average SU Queue Curve 28
5.4 Average Delay Proxy Curve 29
5.5 Secondary User Access Metrics Curve 30
5.6 Primary Protection Metrics Curve 31
5.7 Worst Case Active PU SINR Curve 31
5.8 Idle Sensing Fraction Curve 32
5.9 PU Outage Rate Curve 33
5.10 SU Spectrum Efficiency Curve 33
5.11 PU Protection Reliability Curve 34
5.12 PU SINR Margin Curve 34
5.13 Cumulative Average Reward Curve 35
5.14 Cumulative Average SU Throughput Curve 35
5.15 Epsilon Decay Curve 36
5.16 Reward Convergence Zoom Curve 36
5.17 Throughput vs PU Outage Tradeoff 37
5.18 SU Utilization vs Collision Tradeoff 37
5.19 Queue vs Delay Relationship 38
5.20 Final Performance Summary Bar Chart 38
5.21 Final Protection Collision Summary 39

List of Tables

I Spectrum sensing parameter values and units 18
II 5G NR primary user parameter values and units 19
III SU network parameter values and units 19
IV DQL parameter values and units 20

List of Abbreviations

AP	**A**ccess **P**oint
BAMR	**B**uffer **A**ided **M**ulti-Hop**R**elay
CUE	**C**ellular **U**ser **E**quipment
CR	**C**ognitive **R**adio
CSMA/CA	**C**arrier **S**ensing **M**ultiple **A**cess with **C**ollision **A**voidance
D2D	**D**evice **t**o **D**evice
DSS	**D**ynamic **S**pectrum **S**haring
DQL	**D**eep **Q** **L**earning
DQN	**D**eep **Q** **N**etwork
DSA	**D**ynamic **S**pectrum **A**llocation
FSA	**F**ixed **S**pectrum **A**llocation
FAMA	**F**loor **A**quistion **M**ultiple **A**ccess
5G	**5th** **G**eneration
IOT	**I**nternet **o**f **T**hings
ICSA	**I**mproved **C**hannel **S**election **A**lgorithm
LBT	**L**isten **B**efore **T**alk
LSA	**L**icensed **S**pectrum **A**ccess
LTE	**L**ong - **T**erm **E**voluion
LOS	**L**ine **O**f **S**ight
mmWave	**M**illimeter **W**ave
MAC	**M**edia **A**cess **C**ontrol
NLOS	**N**on **L**ine **O**f **S**ight
OFDM	**O**rthogonal **F**requency **D**ivision **M**ultiplexing
PD	**P**robability of **D**etection
PU	**P**rimary **U**sers
PPO	**P**roximal **P**olicy **O**ptimization
PFA	**P**robability of **F**alse **A**larm
RL	**R**einforcement **L**earning
RAN	**R**adio **A**ccess **N**etwork
SU	**S**econdary **U**sers
SINR	**S**ignal to **I**nterference plus **N**oise **R**atio
TDSC	**T**ime **D**omain **S**ymbol **C**ross-correlation
WI-FI	**W**ireless **F**idelity
WiMAX	**W**orldwide **i**nteroperbility for **M**icrowave **A**cess

Chapter 1

Introduction

1.1. Background

A key concept in improving the effectiveness of spectrum utilization is spectrum sharing. In the past, Traditional fixed spectrum allocation were used by setting specific frequency ranges to a specific services such as cellular networks, WI-FI and television broadcasts. While FSA although is applied to reduce interference among users, its inflexibity and lack of adaptiveness to network traffic resulted in inefficiencies such as lac of adaptability to network traffic [1]. In order to evolve from the static spectrum allocation, DSS was implemented to adapt to the fluctuating demand in traffic. Resources are allocated based on current network needs and user demands such as during peak hours. When network traffic is high, resources might be temporarily shifted from less congested areas to support a surge in data usage.

DSS dynamically adjusts spectrum access and transmission policies to give opportunistic access to SUs when PUs are not utilizing the frequency band thereby, optimizing performance techniques such as system throughput, delay, spectral effficiency and quality of service [2]. particularly for DSS in 5G network, CR is a promising technology that has the potential to revolutionize wireless communication particularly for DSS in 5G wireless networks. CRs give SUs the ability to perceive the licensed environment, identify gaps and modify the transmission behavior to intelligently use spectrum bands licensed to PUs but currently underutilized.

Because opportunistic spectrum access requires precise and low-power detections, The SUs must take wise decisions and optimize throughput while protecting the PUs by watching out the dynamic nature of the primary traffic. And all of this must be accomplished in the face of interference, noise and collisions. Additionally, the development of wireless network intelligence algorithms is compatible with the use of DRL for spectrum management as a result, well designed DRL systems can optimize wireless networks. By integrating Q-Learning with deep neural networks to handle state spaces, DQL expands the use of RL to realistic spectrum sharing between numerous channels and end users.

1.2. Problem Statement

Efficient spectrum usage has become a significant challenge in wireless communication systems due to the growing demand for wireless access. SUs have access to temporal spectrum opportunities because, although licensed spectrum is allocated to PUs, it is not always fully occupied at all times. Permitting SUs to access licensed spectrum is challenging because the quality of service for licensed primary network must be constantly protected.

Fixed sensing thresholds or straightforward LBT methods. are frequently the foundation of conventional spectrum sharing strategies. Although the methods can offer basic interference avoidance, they are insufficiently flexible for dynamic 5G NR situations where SUs compete for scarce spectrum opportunities and PU traffic fluctuates over time. A trade-off between false alarms and missed detection may also result from fixed sensing threshold. The SU may not identify weak PU activity if the threshold is set too high leading to interference. An idle channel may be considered as occupied if the threshold is set too low, wasting spectrum opportunities and lowering seconding throughput.

Therefore, the main issue this thesis addresses is how to secure PU's priority while enabling CR-based SUs to recognize and take advantage of idle spectrum possibilities in a licensed 5GNR n40 band. This thesis suggests a DQL-based DSS framework that integrates adaptive SU access decisions, Markov-based primary traffic modeling, and multi-threshold energy detection to solve this issue. The framework seeks to reduce PU outages, harmful interference, collisions, queue length and latency while increasing SU throughput and spectrum usage.

1.3. Research Objectives

The general objective of this thesis is to design and evaluate a DQL based DSS framework for coexistence between licensed 5G NR PUs and CR-based SUs. The specific objectives are to model 5G NR n40 primary traffic using a four state Markov chain, implement multi-threshold energy detection for secondary sensing, design DQL state and action spaces for channel access, model indoor and outdoor to indoor propagation effects, and evaluate the system using PU protection, SU throughput, DQL reward, collision, utilization, queue, and delay metrics.

1.4. Research Questions

This thesis is guided by four research questions.

- how can spectrum gaps in a licensed 5G NR n40 channel be detected reliably by CR based WI-FI inspired SUs?
- how can a DQL controller learn adaptive channel access decisions from sensing states and traffic behavior?
- how can secondary throughput and utilization be improved without violating the protection requirement of licensed PUs?
- how do interference events, SU collisions, and queue delay change as the DQL agent learns?

1.5. Contribution

The first contribution is the development of a slot level 5G NR and CR-based WI-FI inspired coexistence model using licensed n40 primary operation and opportunistic secondary access. The second contribution is the application of multi-threshold energy sensing, which provides more flexible channel classification than binary idle/busy sensing. The third contribution is a DQL controller that jointly considers sensing state, SU queue, previous action, backoff, and Markov traffic load. The fourth contribution is a performance evaluation framework that combines PU protection and SU access metrics.

1.6. Motivation of the Study and Scope

The specific motivation for this thesis arises from the intersection of two trends that are simultaneously creating new opportunities and new technical challenges in spectrum management. The first is the global deployment of 5G NR systems, which is expanding the population of licensed PUs operating across a wide range of frequency bands, The second is the growing deployment of WI-FI devices in indoor environments. In many deployment situations, the coexistence of 5G NR and WI-FI devices in a shared spectrum is an operational reality rather than a theoretical worry. The interference dynamics must be managed, The effectiveness of spectrum use in this heavy settings heavily relies on how well each network users cooperates and respects the each other's priority.

The scope of the thesis is bounded by a number of specific assumptions that provide a controlled analysis of the proposed framework. The study focuses in an indooor setting for the secondary network, whereas, the primary network comprises of four licensed PUs operating on 4 shared channels. And The secondary network consists of one access point and ten SUs. Instead of being depicting the typical WI-FI operating in licensed bands, the secondary network is represented as CR-based WI-FI inspired network meaning that the users adopt IEEE 802.11ax inspired MAC behaviors such as carrier sensing, CSMA/CA, backoff and TXOP-based access. The DQL agent is constructed as a centralized controller at the SU access point. These boundary conditions are selected to allow a precise and controlled assessment of the fundamental mechanics of the suggested framework; extensions to more complicated scenarios are noted as potential avenues for further research.

1.7. Thesis Organization

Further details of later the chapters are organized as follows; Chapter Two presents relevant review of literature. Chapter three presents the system modeling and methodology. Chapter Four explains the DQL-based spectrum sharing simulation Framework and mathematical formulation. Chapter Four explains the simulation design and parameters. Chapter Five discusses the results. Chapter Six concludes the thesis, points out limitations of the thesis and proposes future work.

Chapter 2

Literature Review

2.1. Introduction

This section provides a review of relevant literature and major concepts related to Dynamic spectrum sharing between 5G NR PUs and CR-based SUs. The review begins with Cognitive radios and spectrum access following that, it examines deep reinforcement learning methods. It also reviews MAC-layer methods such as LBT, contention-based access in heterogeneous networks. Finally propagation loss and system performance studies are addressed

2.1.1. Cognitive Radio Networks for Opportunistic Spectrum Access

The global approach to spectrum management has shifted from FSA to DSA. CR uses DSA to tackle spectrum shortages in a wireless network. PUs allow SUs to access their channel when not using it. And a spectrum handoff takes place when the PUs want their channel back, [3] applied ICSA because it sequentially considers both channel occupancy and signal quality criteria to select a backup channel for a spectrum handoff. A handoff management technique known as NPHS takes place when the system doesn't anticipate for handoff between access points or base stations in advance; it only considers channel occupancy. Thus, ICSA provide a better handoff management by reducing spectrum handoffs 18% when compared with NPHS. [2] addressed spectrum sharing with cooperative sensing technique. The radio identifies spectrum holes when the PUs are not actively using it by performing cooperative sensing enabling many devices to work jointly while this improves accuracy it introduces flaws such as data tampering.

[4] presented a TDSC-based spectrum sensing approach by developing an adaptive CR OFDM system for spectrum sensing.In the study WIMAX and LTE systems were used as a real world examples to illustrate the effectiveness of the TDSC technique.And suggested that an accumulated cross-correlation is necessary for spectrum sensing in CR systems and found the technique has determined whether the frequency band is occupied by PUs even at low SNR. The research [5] attempts to enhance performance metrics such as data rate, outage probability, throughput and delay. The issue the study focuses on is delay because it allows simultaneous transmission and reception across multiple network hop. BAMR has been found to be effective in addressing latency issues. BAMR scheme aims to optimize relay selection in order to increase network reliability and efficiency.

[6] proposed a probabilstic algorithm. The DSS probabilistic approach begins with a SU searching for a channel that goes well with its requirements. Then if the number of handoff surpasses the predefined threshold, the system gives priority to the channel that has higher priority of being idle otherwise using methods like energy detection, the SU senses each channel and updates the probability matrix then determines whether the channel is idle . If the channel is idle, the SU is assigned if not the process is resumed and the probability for the channel is set to zero. This algorithm exemplifies key features of CRs in DSS which are dynamic adaptation to traffic, interference minimization, and effective utilization of idle spectrum while keeping PU's priority secure.

2.1.2. Deep Reinforcement Learning Approaches and Resource Management

Rigid thresholds and preset rules that are set during system design are the foundation of conventional spectrum sensing and access techniques. Everything goes well as long as the operational environment complies with the design assumptions. The problem arises when things don't go as planned as it happens frequently in actual wireless setups. Because of this, they are unable to employ regularities that characterize true spectrum settings or change their behavior in reaction to observable findings. A radio resource management is presented in [7], Allocation approaches were applied to optimize system performance metrics such as system throughput, and quality of service. They discussed radio resource management strategies based on resource allocation. Resources are allocated based on the network's current needs as well as user requests such as during peak hours. During high traffic,resources may be temporarily relocated from less congested areas to handle increase in data demand, One prominent strategy being studied to prevent spectrum under utilization is DSS. According to [1] traditional wireless communication networks often allocate fixed spectrum bands to specific users or services. This technique may lead to inefficiencies like underused spectrum and excessive power utilization. Radio spectrum sharing is introduced as a way to switch from traditional static frequency allocation alignment strategy to dynamic spectrum management.

Prior to the deep learning era, Q-Learning was used to dynamically access spectrum in a series of operations that laid the conceptual ground work for later DQN-based methods. Deep RL spectrum management complements wireless network intelligence. It can provide dependable and efficient learning methods for deployment of controllers, Thus, rather than only resolving a specific coexistence problem, the DQL-based architecture developed in this thesis adds to a more broad vision of intelligent, self optimizing wireless network. A more realistic study [8] examined uncoordinated coexistence between LTE and WI-FI systems using experience replay-based mechanism that gave priority to transactions involving primary interference events during training. The results demonstrated throughput improvements of 19.1% and fairness improvements of 16.4% compared with rule based coexistence schemes, validating the practical utility of Q-learning even in relatively complex coexistence scenarios. The authors noted that the experience replay mechanism was critical to the observed improvements, a finding that influenced the replay buffer design in the present study.

[9] reported a throughput increase from 25% to 48.7% under different user arrival rates by applying simulation-based double DQN. The use this method splits action selection and assessment in to two networks to lessen the overestimation bias that exists in the standard Q-learning.

The researchers in [10] also discovered the dimensionality limits of tabular q-learning can be overcome by deep neural networks capacity to approximate functions, This lead a wave of DRL applications in spectrum management. [11] was one of the researchers to use DQN specifically for cooperative spectrum sensing in CRNs with high dimensional state spaces. They demonstrated that the network could learn effective sensing policies in multi-user scenarios where state space was too large for tabular methods.

[12] States that the PFA of SUs and the probability of PU detection using FAMA-based DQL can be used to determine whether PUs are present in a rayleigh channel. It used Mayfly algorithm for weight optimization instead of using a pre-established rules to dictate behavior;RL enables a system to learn the behavior of its environment by enabling action selection from action space and receiving a reward that encodes the desirability of the outcomes. Over many iterations, the agent converges toward a policy that maximizes cumulative reward, effectively learning, from experience, how to behave optimally in a complex and uncertain environment. [13] presented a hybrid spectrum access strategy that combines orthogonal and non orthogonal access methods for D2D communication in underlay cellular networks. It uses the dynamic time division duplexing network concept where time slots are dynamically assigned to users. A DRL-based system was developed to allow D2D pairings to learn the best spectrum access strategies.

The study [14] presents a thorough analysis of macine learning applications for enhancing IEEE 802.11 performance, learning based techniques helped WI-FI systems to adapt to a variety of network conditions through experience-driven parameter adjustment. Their report claims that machine learning can enhance WI-FI performance in several areas such as channel selection, rate selection, power control and access parameter adaption. The paper pointed out that while coordinated spectrum sharing involves exchanging traffic rate data between coexisting systems, it has the potential to increase overall efficiency, it also adds overhead that needs to be properly managed.

[15] compared policy gradient and q-learning approaches to spectrum access in unlicensed wireless network. The study offered a distributed PPO architecture for contention-based spectrum access, demonstrating that PPO performed well in terms of fair resource distribution among competing nodes and persistent convergence. Their followup study [16] combined contention-based spectrum access with adaptive modulation through deep RL, and showed that DQL and PPO based approaches outperformed adaptive energy detection in both throughput and fairness when operating in unlicensed spectrum.

2.1.3. 5G NR and WI-FI Coexistence

MAC-Layer Coexistence Mechanisms

The coexistence oF 5G NR and WI-FI systems in shared spectrum has been thoroughly studied from the perspective of MAC layer protocol design, particularly the LBT mechanism that both technologies employ. The primary problem is that the LBT protocols used by 5G NR and WI-FI have different contention window sizes, sensing thresholds and TXOP durations. [17] proposed a spatial reuse strategy that combines LBT-based channel sensing with null space projected digital precoding. When compared to the traditional LBT only coexistence the simulation they carried out showed that concurrent transmissions were allowed access co-channel systems, lowering latency and significantly increasing throughput. [18] studied network slicing strategies relevant to 5G MAC and physical layer management. Showing that virtualized network allow for more adaptable interference control and resource allocation. The open RAN-based spectrum sharing that is envisioned

in future enhancements, where DQL controllers operating inside the ORAN architecture might handle spectrum sharing across different slices and technologies simultaneously is contextualized by their work.

[19] combined group of bargaining frameworks with Nash bargaining solutions to offer a fair and efficient spectrum distribution technique for LTE and WI-FI coexistence. The result demonstrated the potential of game theoretic methods as supplements to RL-based strategies. The degree to which the DQL agent's learned policy approaches threshold is an interesting subject for additional research. The Nash bargaining framework offers a helpful benchmark for accessing the fairness of the DQL learned policy. [20] suggested that an adaptive threshold management approach to spectrum sharing between WiMAX and WI-FI networks, allows spectrum sharing between technologies based on dynamically adjusted demand thresholds can greatly increase overall network efficiency. Although this work predates the DRL period and employs simpler approaches than those explored here, it sets an important precedent for threshold-based adaptive spectrum management and provides a useful benchmark for evaluating the additional improvements available through learning-based techniques.

[21] analyzed the effects of MAC parameter selection on collision probability and delay in NR U and WI-FI coexistence situations. Their work showed that the choice of contention window boundaries, sensing thresholds, and TxOP lengths has large and subtle impact on coexistence fairness using a markovian traffic generation model. In particular, they found that NR U systems configured with single network; optimized parameters frequently out competed WI-FI performance and claiming a disproportionate percentage of channel time. The study concluded that achieving fair coexistence requires joint parameter optimization across both technologies, a finding that has direct implications for the reward function design in this thesis.

2.1.4. Propagation Modeling

Indoor and Outdoor to Indoor Propagation

Pathloss is influenced by number of factors sch s distance, buildings furniture. These elements interact in ways that can be challenging. The 3GPP specifications give standardized path loss models to account for the most significant of these factors while remaining manageable enough to be employed in large-scale simulations. In a comparative study of indoor propagation at 3 GHZ and 28 GHZ [5] found that path loss at 28 GHZ is very much higher than at 3 GHZ in both LOS and NLOS conditions. The finding emphasizes the difficulty of indoor coverage at mmWave frequencies, even little obstacles can result in high signal attenuation, thereby; thus stressing the significance of frequency-appropriate propagation models.

[22] performed outdoor-indoor channel experiments at 3.5 GHZ, 4.9 GHZ and 25 GHZ to measure pathloss variables, shadowing fading, and time dispersion characteristics across 5G mid band frequencies. Their study showed correlation between path loss and dispersion parameters. The choice between LOS and NLOS mode is important for simulation accuracy. The higher path loss settings generally exhibit more delayed speed. while [23] study of propagation analysis demonstrated that single wide beam antenna configurations resulted in lower path loss and reduced delay speed.

Antenna Effects and Beamforming

[24] Larger antenna arrays boost data rates and lower latency through more directed beams, but it is at the cost of challenging handover when mobile users move between beam coverage areas . According their findings beam management is an important component of 5G NR performance that needs to be carefully adjusted to the the deployment circumstances. [25] developed a 2x2 antenna array with four ports that targets mid-band frequencies for 5G network users. even if the research focuses on the device rather than the network side of the 5G link, it offers relevant information for understanding antenna capabilities of the SU stations analyzed in the simulation which are assumed to have 0 dbi omini-directional antennas in line with standard WI-FI devices.

[26] presented that massive MIMO antenna arrays can facilitate effective spatial coexistence between cellular and WI-FI networks operating in the same frequency band by utilizing the spatial degrees of freedom offered by large antenna arrays to divert transmit energy away from co-channel WI-FI receivers. Simulation tests showed that large MIMO greatly reduced inter-system interference, allowing simultaneous transmissions without sacrificing system performance. The 1X8 gNB antenna configuration employed in this thesis gives a modest degree of spatial sensitivity appropriate for the 5G deployment.

mmWave and High Frequency DSS

The application of DSS to mm-Wave frequencies present a significant technological challenge as well as opportunity. Rapid channel variation characteristic of mmWave propagation, there was severe path loss, susceptibility to blockage, and rapid channel variations of mmWave propagation. [27] shows weighted compressive sensing recovering strategies for mmWave spectrum sensing were introduced. These techniques use the block sparse structure of mmWave channel metrics to reduce the measurement overhead required for spectrum occupancy detection.

Chapter 3

System Modeling

3.1. Introduction

This chapter presents the methodology the simulation metrics performed followed by the key formulations, energy detection theory, spectrum sensing configuration, markov's chain modeling and the reinforcement learning fundamentals, DQL architecture.

3.2. Methodology

Figure 3.1 depicts the basic methods used in this thesis. The process begins with the formulation of the problem statement and a review of relevant literature, followed by the creation of the system model and channel modeling. Then the PU traffic is modeled and the DQL-based simulation environment is created in matlab. Following that, spectrum sensing and DQL-based decision-making are used to identify the appropriate SU access activities. The SU sends or waits depending on he channel access decisions. Finally, the learning process is updated and performance metrics such as spectrum utilization, PU protection, collisions, delay and queue are accessed.

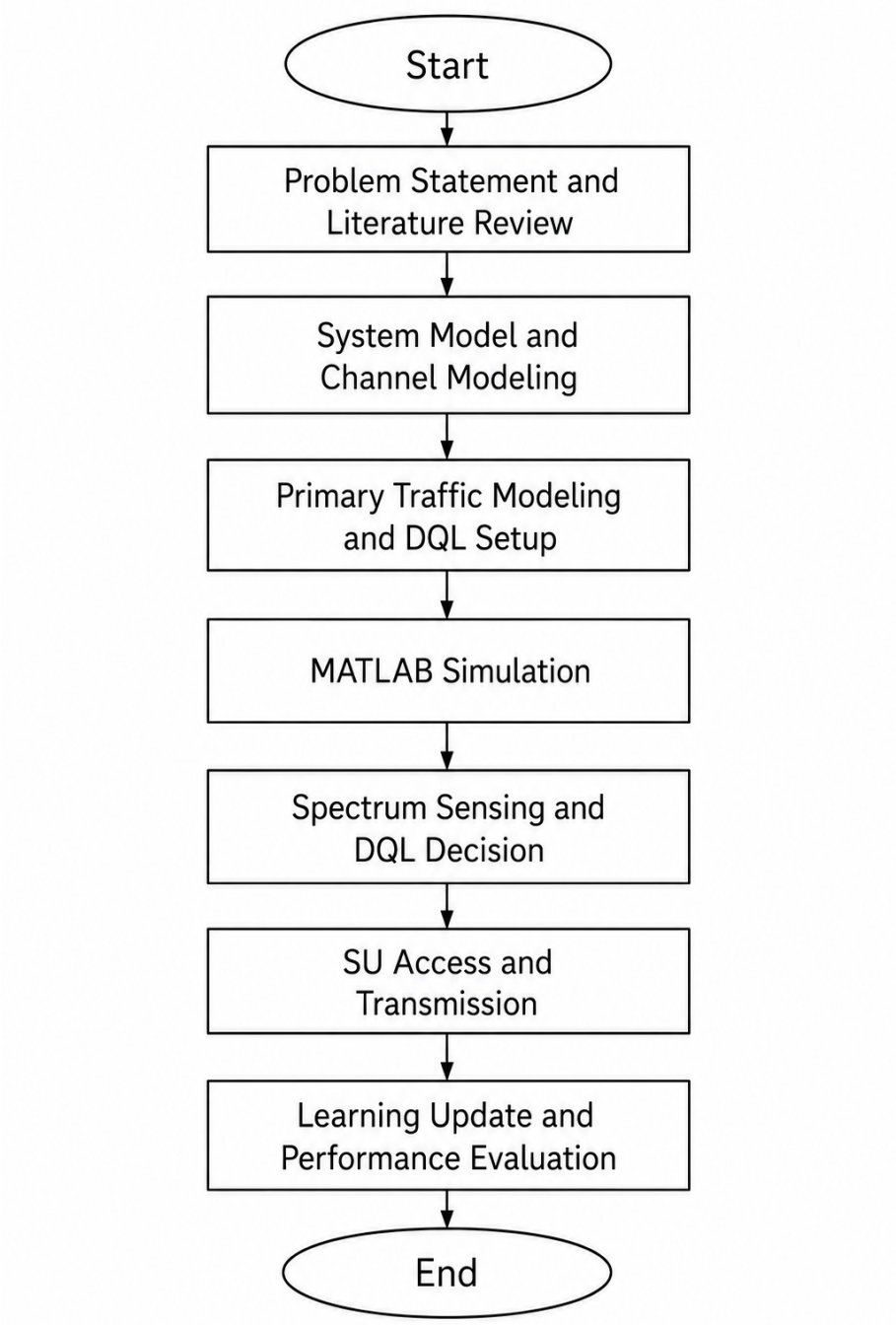


Figure 3.1: Flow Chart

3.3. System Modeling

In the proposed system model, a cognitive radio-based secondary network that mimics WI-FI and a licensed 5G NR primary network exchange spectrum dynamically, The 5G NR network serves as a PU network in the licensed n41 band, while secondary network seeks to exploit unused spectrum without placing the PUs in danger. The PU activity is modeled using a four-state Markov chain traffic generator that represents idle, light, medium and high traffic conditions. Spectrum sensing is used by the secondary access point to view shared spectrum, and access decisions are managed by DSS controller based DQL.

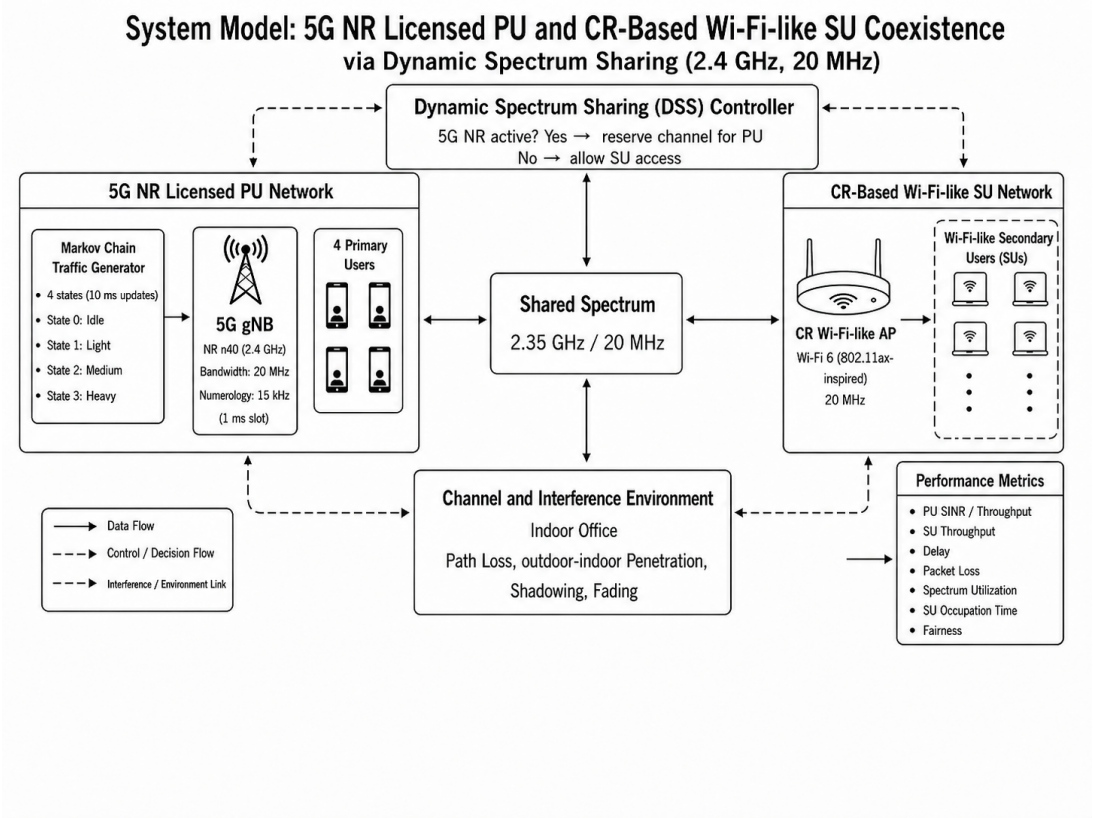


Figure 3.2: System Model

3.3.1. Energy Detection Theory

Energy detection is most widely used spectrum sensing techniques. It is not necessary to know the major signal's structure in advance. The sensing device simply computes the total signal energy received in the shared channel and compares it to a threshold. The AP uses energy detection to sense if the channel is idle. The received energy on channel k is estimated over N samples as

$$E_k = \frac{1}{N} \sum_{n=1}^{N} |r_k(n)|^2 \tag{1}$$

where E_k is the energy measured k, N is the number of sensing samples, and $r_k(n)$ is the received signal sample.

The measured energy is also normalized by the noise floor power N_0 as

$$E_{k,\text{norm}} = \frac{E_k}{N_0} \tag{2}$$

The state of the channel is then computed from :

$$S_k = \begin{cases} 0, & E_{k,\text{norm}} < \lambda_1, \quad \text{Idle channel} \\ 1, & \lambda_1 \leq E_{k,\text{norm}} < \lambda_2, \quad \text{Uncertain channel} \\ 2, & \lambda_2 \leq E_{k,\text{norm}} < \lambda_3, \quad \text{PU likely present} \\ 3, & E_{k,\text{norm}} \geq \lambda_3, \quad \text{PU active} \end{cases} \tag{3}$$

In this thesis, the thresholds are:

$$\lambda_1 = 1.1, \qquad \lambda_2 = 1.4, \qquad \lambda_3 = 1.6 \tag{4}$$

where S_k denotes the sensing state of channel k. If $S_k = 0$, the channel is regarded as idle and the SU is free to transmit. If $S_k = 1$, the channel is not sure to be free andreduced power is used. If $S_k = 2$, the channel is likely being used and resensing is favored. If $S_k = 3$, PU is utilizing the channel and SU is restrained from transmitting.

3.3.2. The Detection-False Alarm Trade-off

The probability of faalse alarm is an important sensing metric from the perspective of SU. It occurs when PU is not present but the SU considers the channel as occupied. In this case, the SU decides not to transmit even though the spectrum is available causing secondary throughput to decrease and resulting in missed spectrum opportunities. Therefore, even while a high probability of detection is necessary to protect PU from interference, an excessively high PFA reduces the efficiency of opportunistic spectrum access.

In the simulation framework used in this thesis, the sensing interval is fixed at 1 ms. And the thresholds are set to lambda values of 1.1, 1.4, and 1.6,corresponding to the three classification categories described above. The DQL agent is exposed to the sensing errors that arise from these threshold choices; both false alarms that cause it to classify occupied channels as idle and miss detections that cause it to transmit on occupied channels and it learns a policy that is robust to these errors through experience, adjusting its access decisions based on the observed consequences of past choices.

3.3.3. Reinforcement Learning Fundamentals

RL is a paradigm for learning decision-making policies. RL learns from rewards that show how desirable previous actions were. In order to increase the likelihood of high reward actions, the RL agent repeatedly examines the present state of the environment, chooses action, gets rewarded, moves to a new state and modifies its policy. The agent approaches a strategy that maximizes its predicted cumulative reward over numerous iterations. The markov decision process and the markov PU traffic state serve as the one of the formal foundation for RL.

3.3.4. Markov Chain Traffic Modeling

The dynamic of PU activity, the statistical mechanism by which the channel shifts from busy to idle , is a crucial component in determining the effective spectrum sharing process.

A secondary device that can accurately predict primary activity would achieve higher throughput and lower interference rates than one that views primary transitions as entirely random events. It is challenging to model these dynamics in a realistic and manageable way. Markov chain models have been widely employed for this purpose in the CRN literature because they represent the essential feature of channel dynamics, the presence of future state on existing state, in a mathematically manageable way.

By differentiating between states that can be read as idle;state 0, light load; state 1 , medium load; state 2 and heavy load state 3 the markov chain used in this simulation expands up on the two state model representing a larger collection of fundamental traffic behaviors. These states are characterized primary activity which establishes the quality of spectral opportunity as well as minimized interference associated with SU transmission. In order to ensure that the DQL agent experiences a variety of patterns during training, state transitions take place every 10ms that is every 10 slots, with heterogeneous transition matrices across the four channels.

Discrete Markov's chain modes PU traffic as follows. Let X_t denote the PU traffic state at time slot t, where

$$X_t \in \{0, 1, 2, 3\} \tag{5}$$

The four states represent idle, light, medium, and heavy traffic conditions, respectively. The Markov property is expressed as

$$\Pr(X_{t+1} = j \mid X_t = i, X_{t-1}, \ldots, X_0) = \Pr(X_{t+1} = j \mid X_t = i) \tag{6}$$

This means that the next PU traffic state depends only on the current state. The transition probability from state i to state j is defined as

$$P_{ij} = \Pr(X_{t+1} = j \mid X_t = i) \tag{7}$$

The transition probability matrix applied in this thesis is

$$\mathbf{P} = \begin{bmatrix} 0.70 & 0.20 & 0.08 & 0.02 \\ 0.25 & 0.50 & 0.20 & 0.05 \\ 0.10 & 0.25 & 0.50 & 0.15 \\ 0.05 & 0.15 & 0.30 & 0.50 \end{bmatrix} \tag{8}$$

where each row satisfies

$$\sum_{j=0}^{3} P_{ij} = 1, \qquad 0 \leq P_{ij} \leq 1 \tag{9}$$

The number of active PUs is determined from the current traffic state as

$$N_{\text{PU,active}}(X_t) = \begin{cases} 0, & X_t = 0 \quad \text{Idle} \\ 1, & X_t = 1 \quad \text{Light traffic} \\ 2, & X_t = 2 \quad \text{Medium traffic} \\ 4, & X_t = 3 \quad \text{Heavy traffic} \end{cases} \tag{10}$$

The simulation uses a slot duration of $T_s = 1\,\text{ms}$, while the Markov traffic state is updated every ten slots. Therefore,

$$T_M = 10T_s = 10\,\text{ms} \tag{11}$$

Let the state probability vector be

$$\boldsymbol{\pi}_t = \begin{bmatrix}\pi_0(t) & \pi_1(t) & \pi_2(t) & \pi_3(t)\end{bmatrix} \tag{12}$$

Then the state probability evolves according to

$$\boldsymbol{\pi}_{t+1} = \boldsymbol{\pi}_t \mathbf{P} \tag{13}$$

For an n-step transition, the state probability is

$$\boldsymbol{\pi}_{t+n} = \boldsymbol{\pi}_t \mathbf{P}^n \tag{14}$$

The stationary distribution of the Markov chain is given by

$$\boldsymbol{\pi}^* = \boldsymbol{\pi}^* \mathbf{P} \tag{15}$$

subject to

$$\sum_{i=0}^{3} \pi_i^* = 1, \qquad \pi_i^* \geq 0 \tag{16}$$

3.3.5. Deep Q-Learning Formulation

The proposed dynamic spectrum sharing problem is formulated as a DQL problem in which the secondary access point acts as the learning agent. At each time slot t, the agent observes the radio environment, selects a spectrum access action, receives a reward, and updates its policy to maximize long term cumulative reward.

3.3.6. State Space

The DQL state vector is defined using the sensing results of the K shared channels, the SU queue length, the previous action, the backoff counter, and the Markov primary traffic load state. The state vector is expressed as

$$\mathbf{s}_t = [S_1(t), S_2(t), \ldots, S_K(t), Q_t, a_{t-1}, B_t, X_t]^T \tag{17}$$

where $S_k(t)$ is the sensing state of channel k, Q_t is the SU queue length, a_{t-1} is the previous action, B_t is the backoff counter, and X_t is the Markov traffic state of the primary network. Since the simulation uses $K = 4$ shared channels, the state vector becomes

$$\mathbf{s}_t = [S_1(t), S_2(t), S_3(t), S_4(t), Q_t, a_{t-1}, B_t, X_t]^T \tag{18}$$

The sensing state of each channel is represented as

$$S_k(t) \in \{0, 1, 2, 3\} \tag{19}$$

where 0 indicates an idle channel, 1 indicates an uncertain channel, 2 indicates that a PU is likely present, and 3 indicates an active PU channel.

3.3.7. Action Space

The DQL action consists of selecting one of the K shared channels and one of the available access modes. The action is defined as

$$a_t = (k, m) \tag{20}$$

where

$$k \in \{1, 2, \ldots, K\} \tag{21}$$

and

$$m \in \{\text{TX-full}, \text{TX-reduced}, \text{Re-sense}, \text{Defer}, \text{Silent}\}. \tag{22}$$

Therefore, the total number of possible DQL actions is

$$|\mathcal{A}| = K \times M \tag{23}$$

where $M = 5$ is the number of access modes. For $K = 4$ shared channels,

$$|\mathcal{A}| = 4 \times 5 = 20. \tag{24}$$

3.3.8. Q-Value Function

The objective of the DQL agent is to learn the optimal action value function, which estimates the expected discounted cumulative reward obtained by taking action a_t in state $\mathbf{s}_t$ and following the learned policy afterward:

$$Q^{\pi}(\mathbf{s}_t, a_t) = \mathbb{E}_{\pi}\left[\sum_{\tau=0}^{\infty} \gamma^{\tau} r_{t+\tau} \mid \mathbf{s}_t, a_t\right] \tag{25}$$

where r_t is the immediate reward and γ is the discount factor. In this simulation,

$$\gamma = 0.9. \tag{26}$$

The optimal Q-value function satisfies the Bellman optimality equation:

$$Q^{*}(\mathbf{s}_t, a_t) = \mathbb{E}\left[r_t + \gamma \max_{a'} Q^{*}(\mathbf{s}_{t+1}, a')\right]. \tag{27}$$

3.3.9. Deep Q-Networks

Standard tabular Q-learning stores a separate Q-value estimate for each state-action pair in a table. This approach is tractable for problems with small, discrete state spaces but breaks down when the state space is large, continuous, or high dimensional as is the case

in practical spectrum access problems where the state includes continuous valued energy measurements across multiple channels and multiple time steps. In DQL, the Q-value function is approximated using a neural network with parameters θ:

$$Q(\mathbf{s}_t, a_t; \theta) \approx Q^*(\mathbf{s}_t, a_t). \tag{28}$$

The network takes the state vector $\mathbf{s}_t$ as input and produces one Q-value for each possible action:

$$\mathbf{q}_t = \left[Q(\mathbf{s}_t, a_1; \theta), Q(\mathbf{s}_t, a_2; \theta), \ldots, Q(\mathbf{s}_t, a_{|\mathcal{A}|}; \theta)\right]^T. \tag{29}$$

3.3.10. DQL Target Value

The DQL target value is calculated using the target network with parameters θ^-:

$$y_i = r_i + \gamma \max_{a'} Q(\mathbf{s}_{i+1}, a'; \theta^-) \tag{30}$$

where θ^- represents the target network parameters. The use of a target network improves training stability by reducing rapid changes in the target value.

3.3.11. Loss Function

The DQL loss function is defined as the mean squared error between the target value and the predicted Q-value:

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{i=1}^{M} \left[y_i - Q(\mathbf{s}_i, a_i; \theta)\right]^2. \tag{31}$$

Substituting the target value gives

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{i=1}^{M} \left[r_i + \gamma \max_{a'} Q(\mathbf{s}_{i+1}, a'; \theta^-) - Q(\mathbf{s}_i, a_i; \theta)\right]^2. \tag{32}$$

Reward Function Design

The reward function is the most consequential design choice in any RL-based spectrum sharing system because it directly determines what the agent optimises. A poorly designed reward function can lead to agents that achieve high reward through undesired behaviours for example, an agent that avoids all primary interference by never transmitting, or one that maximizes short-term throughput at the cost of catastrophic primary interference.

The reward function is employed to balance SU's throughput while keeping the PU's priority and protection. positive reward is given for successful PU protection while penality is imposed when SU collision occurs,PU outages,interference with actively using PU and queue-related delay.The DQL agents learns how to select actions that improves SU throughput and effective utilization of vacant spectrums while keeping the PUs protected.

$$R_t = w_T T_{SU}(t) - w_O \mathbb{1}_{\text{outage}}(t) - w_I \mathbb{1}_{\text{interference}}(t) - w_C \mathbb{1}_{\text{collision}}(t) - w_D Q_{SU}(t) \tag{33}$$

$$R_t = 1.0\,T_{SU}(t) - 12.0\,O_{PU}(t) - 3.0\,I_{PU}(t) - 2.0\,C_{SU}(t) - 0.02\,Q_{SU}(t) \tag{34}$$

where R_t is the immediate reward at time slot t, $T_{SU}(t)$ is the secondary-user throughput in Mbps, $O_{PU}(t)$ is the primary-user outage indicator, $I_{PU}(t)$ is the interference event indicator, $C_{SU}(t)$ is the secondary-user collision indicator, and $Q_{SU}(t)$ is the secondary-user queue length in packets.

3.3.12. LBT Protocol

Before launching transmission, a transmitting device must sense the channel and confirm the channel is idle according to the LBT protocol. It is the foundation of CSMA/CA method utilized by WI-FI. In the context of this thesis, LBT serves a complementary role to the DQL controller: while the DQL agent makes strategic decisions about which channel to attempt to access, LBT provides a tactical safeguard that prevents the SU from transmitting if a primary signal is detected in the moments immediately preceding the transmission attempt.

The specific LBT variant employed in this simulation is based on the 802.11ax standard's CSMA/CA mechanism, which includes carrier sense, virtual carrier sense through the Network Allocation Vector, random backoff, and exponential contention window growth. When the SU access point decides to transmit on a given channel, it first performs a clear channel assessment by comparing the measured energy against the sensing threshold. If the channel is assessed as idle, the AP initiates a random backoff countdown by drawing a random integer from [0, CWmin]. The backoff counter decrements by one for each idle slot and is frozen when the channel becomes busy. When the counter reaches zero, the AP transmits a packet of 1500 bytes during the TxOP duration of 2 ms. If a collision occurs either with a primary transmission or with another SU; the AP doubles the contention window;(up to CWmax) and repeats the backoff process. The integration of LBT with DQL creates a two level access control architecture. At the strategic level, the DQL agent selects the channel on which the SU will attempt to access this decision is made once per slot and reflects the agent's learned policy for maximizing long term reward. At the tactical level, LBT provides a real time check that catches cases where the channel has become occupied in the interval between the DQL decision and the actual transmission attempt, a scenario that is particularly relevant in fast varying traffic conditions. Together, these two mechanisms provide both the adaptability of a learning based controller and the reactivity of a real time sensing protocol.

Chapter 4

DQL-based spectrum sharing simulation Framework

The DQL-based spectrum sharing simulation framework created to access opportunistic access between a CR-based SU network and a licensed 5G NR primary network is presented in this section.The framework is built as a system-level, slot-based simulation here secondary access point use DQL to update its learning policy choose access actions. Primary traffic modeling, multithreshold energy detection, indoor and out door-indoor propagation, and spectrum sensing configurations are all included in this simulation.

4.1. Simulation Parameters

The table below represents the system level parameters for the simulation environment.

Table I: Spectrum sensing parameter values and units

Symbol	**Value**	**Description**
N	500 samples	Number of samples used for each sensing test
T_s	1 ms	Spectrum sensing interval
N_0	-100 dBm	Noise floor used for energy detection
λ_1	1.1	First energy threshold for idle classification
λ_2	1.4	Second energy threshold for uncertain/PU-likely classification
λ_3	1.6	Third energy threshold for PU-active classification
σ_n^2	1	Normalized noise variance

Table II: 5G NR primary user parameter values and units

Symbol	**Value**	**Description**
f_c	2.35 GHz	Carrier frequency in the NR n40 band
B	20 MHz	Channel bandwidth
K	4	Number of shared channels
N_{PU}	4	Number of licensed PUs
μ	0	5G NR numerology index
Δf	15 kHz	Subcarrier spacing
T_{slot}	1 ms	Slot duration
P_{gNB}	30 dBm	gNB transmit power
G_{gNB}	8 dBi	gNB antenna gain
$\gamma_{PU,th}$	10 dB	PU SINR outage threshold

Table III: SU network parameter values and units

Symbol	**Value**	**Description**
N_{AP}	1	Number of secondary access points
N_{SU}	10	Number of SUs
$P_{SU,full}$	15 dBm	Full SU transmit power
$P_{SU,red}$	10 dBm	Reduced SU transmit power
G_{SU}	0 dBi	SU antenna gain
CW_{min}	15	Minimum contention window
CW_{max}	1023	Maximum contention window
AIFSN	3	Arbitration inter-frame spacing number
$TXOP_{normal}$	2 ms	Normal transmission opportunity duration
$TXOP_{short}$	1 ms	Short transmission opportunity duration
L_p	1500 bytes	Packet size

Table IV: DQL parameter values and units

Symbol	**Value**	**Description**
α	0.001	Learning rate
γ	0.9	Discount factor
ϵ_{start}	1.0	Initial exploration probability
ϵ_{end}	0.05	Final exploration probability
E_{decay}	400 episodes	Epsilon decay duration
M_r	50000	Replay memory size
B_s	64	Mini-batch size
H_1	64 neurons	First hidden layer size
H_2	64 neurons	Second hidden layer size
$\|\mathcal{A}\|$	20	Total number of DQL actions

Before examining specific performance metrics, it is useful to examine the structural links between the system's components and the signal flows that govern training and realtime operation. The four gNB in the primary 5GNR network in the n40 band to generate down link traffic, according to the four state markov chain model. The secondary CR-based network, which works in an internal office setting, consists of ten stations and one access point. The access point handles all sensing and controller functions, the stations execute the transmissions that AP instructs and return the results to the AP so that they can be incorporated in to the reward signal.

The outdoor to indoor propagation model presented in chapter 4, describes the radio channel that links the primary and secondary networks. At the simulation's reference distance and frequency, the average pathloss from the gNBs to the SU AP, including outdoor pathloss, building penetration, indoor pathloss is roughly 95db. This corresponds to a received SNR at the SU AP in the range of -5 to +5 dB to the gNB transmit power of 30dBm. The multi-threshold technique and the DQL agent's uncertainty management are especially useful because this SNR range puts the system in the intermediate sensing accuracy region, far above the noise wall, but not in the high SNR realm, where energy detection is practically perfect.

The simulation is constructed as an event-driven, discrete-time system with a time-slot percision 1ms. This slot duration was selected to match the 5G NR subframe structure under numerology, where a single subframe, which is the fundamental unit of 5G NR resource allocation, takes up precisely one millisecond. The modeled primary traffic dynamics and the DQL agent's decision timing are guaranteed to be consistent with the temporal structure of the underlying 5G standard when the simulation slot is aligned with 5G NR subframe. The simulation examines the communication between a primary and secondary network that are co-located and share the same licensed frequency band. Both have a 20 MHZ channel bandwidth and operate in the n40 band at 2.35 GHZ. The primary network generates traffic according to a stochastic model calibrated to represent realistic 5G downlink traffic variability. The secondary network monitors the primary's channel, makes access decisions, and transmits data according to the policy learned by the DQL agent.

The training loop of the simulation consists of 500 episodes. The DQL agent's e-greedy exploration parameters is changed in accordance with a decay schedule at the beginning of each episode, and the markov chain states are reset to randomly chosen initial circumstances. PU traffic gets update every 10 slots. AP channel sensing, DQL action selection, LBT contention and transmission, channel propagation and interference computation metric recording are all carried out at each time slot as the simulation moves through each episode.A mini-batch sampled from the experience replay buffer is used to update the DQL network at the end of each episode.

4.1.1. Primary User Traffic Model

The PU activity pattern is represented by by the PU traffic model.four licensed users utilizing the four channels make up the main network in this simulation, and a four-state Markov chain is used to generate the traffic flow. The states correspond to varying levels of principal user channel occupancy and depict idle, light, medium and heavy traffic circumstances.

4.1.2. Secondary User Network Model

The ten SUs and one access point the secondary CR-based network that operates in an indoor office setting. The access point is the secondary network's centralized intelligence. I t manages the DQL controller, performs channel sensing and synchronizes the ten station's transmissions. With the MAC parameters listed in the simulation parameter table, the statins carry out broadcats that are directed by the accesspoint using 802.11ax CSMA/CA.

4.1.3. Path-Loss and Propagation Model

The propagation model considers outdoor gNB transmission, outdoor-to-indoor penetration into the smart-building office, and indoor AP-to-SU/SU-to-SU links. The carrier frequency is set to $f_c = 2.35$ GHz. The outdoor gNB is placed at height $h_{\mathrm{gNB}} = 10$ m, the licensed PUs are placed at $h_{\mathrm{PU}} = 1.5$ m, and the indoor AP and SUs are placed at $h_{\mathrm{AP}} = h_{\mathrm{SU}} = 1$ m.

The three-dimensional distance between a transmitter and receiver is calculated as

$$d_{3D} = \sqrt{d_{2D}^2 + (h_t - h_r)^2} \tag{1}$$

where d_{2D} is the horizontal distance, h_t is the transmitter height, and h_r is the receiver height.

For the outdoor gNB propagation link, a simplified 3GPP UMi-like path-loss model is used:

$$PL_{\mathrm{UMi}}(d_{3D}, f_c) = 32.4 + 21 \log_{10}(d_{3D}) + 20 \log_{10}(f_c) \tag{2}$$

where PL_{UMi} is expressed in dB, d_{3D} is in meters, and f_c is in GHz.

The outdoor-to-indoor path loss from the gNB into the smart-building office is modeled by adding building penetration loss:

$$PL_{\mathrm{O2I}} = PL_{\mathrm{UMi}} + L_{\mathrm{building}} \tag{3}$$

where

$$L_{\mathrm{building}} = 20 \text{ dB}. \tag{4}$$

For indoor AP-to-SU and SU-to-SU line-of-sight links, the indoor office LOS path loss is given by

$$PL_{\mathrm{InH,LOS}}(d_{3D}, f_c) = 32.4 + 17.3 \log_{10}(d_{3D}) + 20 \log_{10}(f_c). \tag{5}$$

For indoor non-line-of-sight links, the indoor office NLOS path loss is modeled as

$$PL_{\mathrm{InH,NLOS}}(d_{3D}, f_c) = \max\left[PL_{\mathrm{InH,LOS}}, 17.3 + 38.3 \log_{10}(d_{3D}) + 24.9 \log_{10}(f_c)\right]. \tag{6}$$

Shadow fading is included as a zero-mean Gaussian random variable:

$$PL_{\mathrm{total}} = PL + X_\sigma \tag{7}$$

where

$$X_\sigma \sim \mathcal{N}(0, \sigma^2). \tag{8}$$

In this simulation, the LOS and NLOS shadow fading standard deviations are

$$\sigma_{\mathrm{LOS}} = 3 \text{ dB}, \qquad \sigma_{\mathrm{NLOS}} = 8 \text{ dB}. \tag{9}$$

Small-scale Rayleigh fading is modeled using a complex Gaussian channel coefficient:

$$h = \frac{1}{\sqrt{2}} (h_I + j h_Q), \qquad h_I, h_Q \sim \mathcal{N}(0, 1). \tag{10}$$

The corresponding fading gain in dB is

$$G_f = 10 \log_{10} \left(|h|^2 \right). \tag{11}$$

The received power is calculated as

$$P_r(\text{dBm}) = P_t(\text{dBm}) + G_t(\text{dBi}) + G_r(\text{dBi}) - PL_{\text{total}}(\text{dB}) + G_f(\text{dB}) \tag{12}$$

where P_t is the transmit power, G_t and G_r are the transmitter and receiver antenna gains, PL_{total} is the total path loss including shadowing and penetration loss when applicable, and G_f is the Rayleigh fading gain.

For the gNB-to-PU desired signal, the received power is

$$P_{gNB \to PU_i} = P_{gNB} + G_{gNB} + G_{PU} - PL_{\text{UMi}}(d_{gNB,PU_i}, f_c) + G_f. \tag{13}$$

For the AP-to-SU indoor link, the received power is

$$P_{AP \to SU_i} = P_{SU} + G_{AP} + G_{SU} - PL_{\text{InH}}(d_{AP,SU_i}, f_c) + G_f. \tag{14}$$

For the SU/AP-to-PU interference link, the received interference power is

$$I_{SU \to PU_i} = P_{SU} + G_{SU} + G_{PU} - PL_{\text{UMi}}(d_{SU,PU_i}, f_c) - L_{\text{building}} + G_f. \tag{15}$$

The SINR experienced by the primary receiver when the SU is transmitting on the same channel is:

$$\text{SINR}_{PU}(\text{dB}) = 10 \log_{10} \left(\frac{P_{\text{rx},PU}}{N_{\text{floor}} + I_{SU}} \right) \tag{16}$$

where SINR_{PU} is the signal-to-interference-plus-noise ratio experienced by the primary receiver, $P_{\text{rx},PU}$ is the received signal power at the primary receiver from its serving gNB, N_{floor} is the receiver noise-floor power -100dbm, and I_{SU} is the interference power generated by the SU at the primary receiver location. A PU outage event is declared whenever SINR_{PU} falls below 10 dB. This threshold represents the minimum SINR required for reliable primary downlink decoding under the assumed modulation and coding scheme. The interference power depends on the SU transmit power; 15 dBm at full power and 10 dBm at reduced power, the path loss between the SU and the primary receiver, and the shadowing realization at the SU-to-PU receiver link. The building penetration loss applies to this link in reverse that is from indoor SU to outdoor primary receiver, contributing to the spatial isolation between the secondary and primary networks that is critical for enabling opportunistic coexistence.

4.1.4. System Component Interactions

The interaction among system components follows a well-defined signal and control flow that determines how information propagates through the system at each simulation slot. Understanding this flow is essential to interpreting both the system's operational behavior and the performance metrics recorded during simulation. Athe the beginning of each slot, the primary gNBs emit downlink signas on their respective channels based on the status of their markov chain traffic models. The SU access point generates a set of energy measurements E(K) for K=1,2,3,4 by using energy detection technique to determine the received signal energy on each of the four channels. These readings are compared to multi-thresholds to establish a channel state for each channel. The agent assesses its current

policy and presents action. If the agent is in exploitation mode, the action with greatest Q-value is chosen and if the agent is in exploration mode, a random action is chosen.

If the selected action is to access a channel, the SU access point initiates the LBT contention process. Inorder to accomplish this, it draws a random backoff counter, performs a carrier sense checkand decreases it throughout the course of the consecutive idle slots until it hits Zero. The access point ten send out a 1500 byte packet during the 2ms TxOP. If the channel becomes busy during the backoff countdown, the countdownis stopped until it returns to idle. The outcome of transmission attempt which can be success collision with PU, or collision with other SU is determined by comparing the received SINR at each affected receiver. After the transmission outcome is determined, the reward signal is computed and the transition is recorded in the experience replay buffer At the end of each episode, a mini-batch is sampled from the replay buffer to update the DQL network weights via bachpropagation. The goal network weights are modifed at predefined intervals to guarantee training stability.

4.1.5. Spectrum Sensing Configuration

Multi-Threshold Classification

The multi-threshold energy detection method classifies each channel in to categories with respect to three thresholds based on the measured energy level. The goal is to provide the DQL agent with a more intricate representation of channel state uncertainty than would be achievable with binary classification. By confidentially differentiating occupied states, the scheme allows the agent to adapt its access aggressiveness to the confidence level of the sensing outcome, thereby, reducing the risk of PU interference without unduly restricting secondary access.

Sensing Error Modeling

The simulation specifically models two forms of sensing errors; noise fluctuations and limited sensing intervals. Thermal noise can introduce random variation in the measured energy around the threshold, causing the classification to diverge from the true channel state. The finite sensing interval limits the energy estimate's precision at 500 samples per 1ms sensing interval at 20 MHZ bandwidth, the energy estimate varies depending on the SNR and channel state. The multi-threshold technique reduces the effective PMD at the cost of a slight increase in the idle spectrum sensing; the fraction of genuinely idle slots in which the SU defers transmission due to uncertain sensing outcome is a trade-off that the DQL agent learns to manage through experience.

Chapter 5

Results and Discussion

5.1. Introduction

The simulation results of the proposed DQL-based DSS framework between licensed 5G NR PUs and CR-based SUs are presented and discussed in this chapter. Matlab is used to run the simulation. The results are used to access how well the secondary network takes advantage of idle spectrum while keeping PUs priority protected Key metrics such as DQL reward convergence, queue length, delay, collision events and PU protection are used to examine the simulation performance.

5.1.1. Simulation Results

The DQL reward increases as shown in Figure 5.1.The DQL reward trajectory is computed over each slots in every episode. Starting at around 35 in the first episode it gradually increases to about 76 to 78 by the end of the training episode. The three distinct trajectory stages demonstrate the agent's interaction with the spectrum environment is expanding. In the first 100 episodes the agent earns experience across the action space, the neural network weight shifts rapidly as the Q-function adapt to the broad range of state-action parings. The reward increases quickly as the agent learns to identify the pattern

In the second stage which is from episodes 100 to 350, the reward also increases more gradually, the agent must now learn to use correlations within each channel's markov process to differentiate between channels that are temporarily idle in between spikes in traffic and channels that are idle due to low load.

The third phase starts at episode 350. Because of the stochastic nature of the traffic model, the reward stablizes at roughly 76 to 78, with minor episode to episode changes.

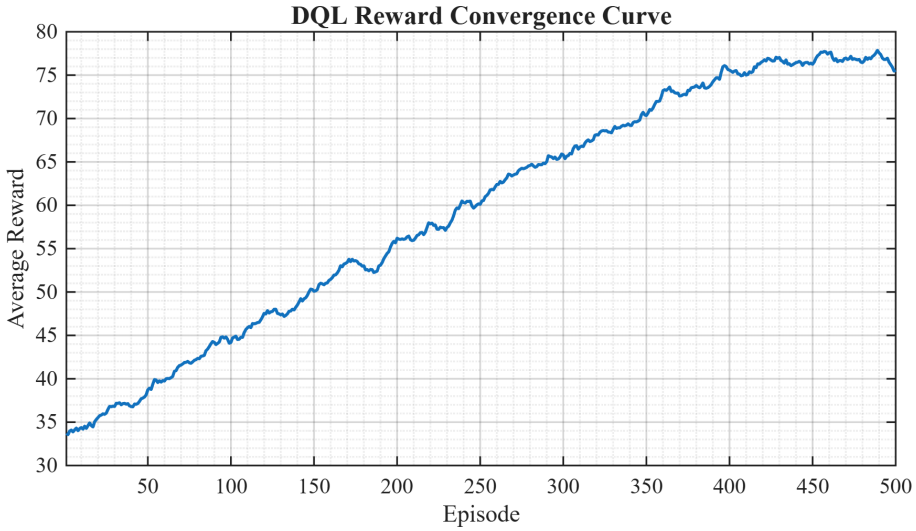


Figure 5.1: DQL Reward Convergence Curve

5.1.2. Secondary User Throughput

Figure 5.2 shows the SU throughput curve. An initial throughput of 36Mbps results from the SU's repeated attempts to access occupied channels and collisions that reduce effective throughput below the theoretical limit. Throughput gradually rises during the learning stages reaching about 80Mbps by episode 500. Without altering the parameters, this results in a throughput improvement of roughly 115-120% over the untrained baseline, which is greater than a doubling of secondary throughput attained only from enhanced channel selection.

Throughout the training phases, the throughput improvement is not linear. It is most gradual during the refinement phase, when the remaining gains necessitate the finer-grained discrimination previously mentioned, and most rapid during the first exploration phase, when the agent rapidly learns to avoid the most highly loaded channels. Throughput has stabilized at roughly 79 to 80 Mbps by the convergence phase. The channel occupancy statistics can be used to estimate the theoretical maximum secondary throughput which is possible if the SU could consistently find and access the channel with the lowest primary occupancy with perfect accuracy. By the end of training, the DQL agent reaches between 75 and 80% of the theoretical maximum, which is comparable to simpler access strategies and offers plenty of opportunity for development through more complex multichannel access rules.

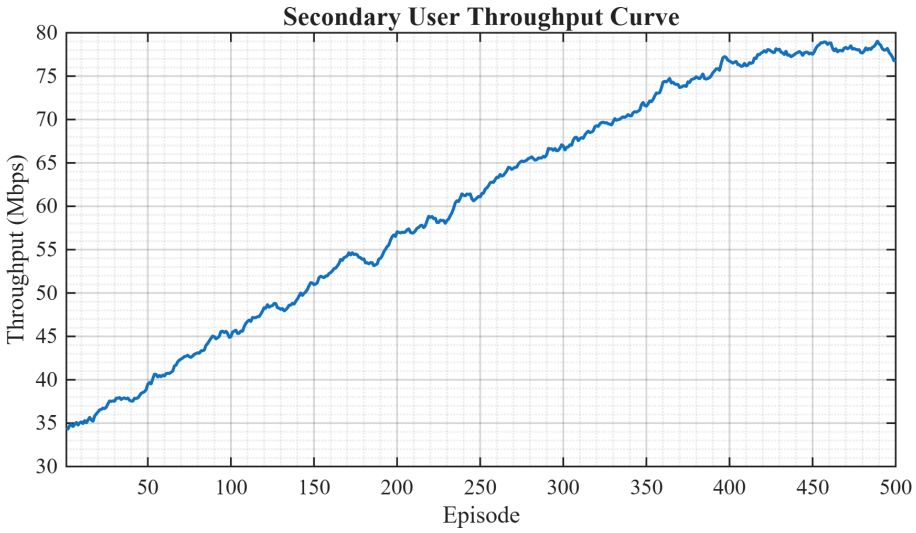


Figure 5.2: Secondary User Throughput Curve

5.1.3. Queue Length and Delay Dynamics

Due to the combination of modest throughput and frequent collision-induced transmissions,in Figure 5.3 the SU queue length ranged from 8 to 10 packets each episode. Each collision requires a full contention window backoff before the next transmission attempt, during which new packets arrive and queue length grows. The backoff overhead drops, throughput rises and queue length lowers as the DQL agent learns to choose channels with lower collision rates. The queue length has decreased to about 2 to 3 packets by episode 300. And by episode 400 it stablizes at about one packet; showing the system is at steady state with amost equal packet arrivals and departures. For secondary applications that are sensitive to delays, this almost negligible queue length is important.

The queue length trajectory is mirrored by delay proxy, as depicted in Figure 5.4, a normalize measure of the number of slots a packet waits before successful transmission that include both queuing time and backoff time. It decreases from 8 to 10 slots in early training time to roughly 0.8 to 1 slot. Since the simulation uses a 1 ms slot, this indicates a substantial reduction in delay.Thus, the delay proxy represents a substantial improvement in the secondary network's ability to support latency sensitive traffic.

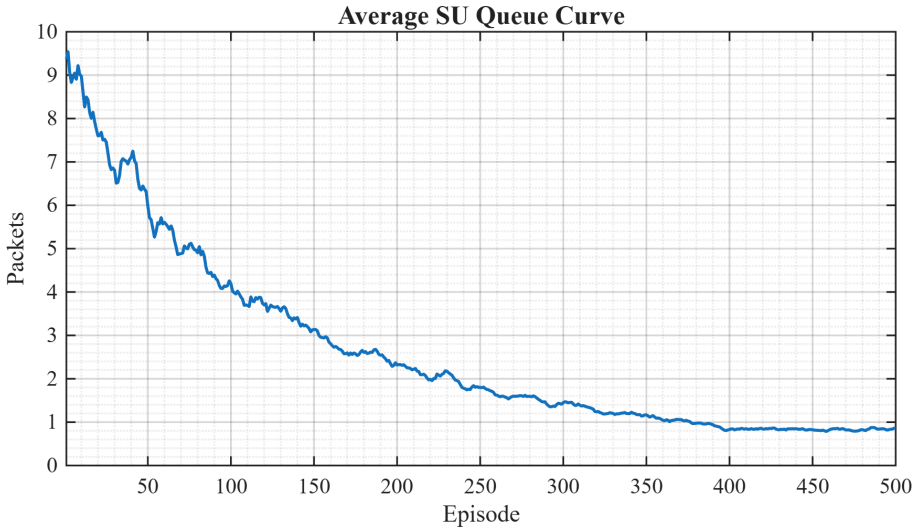


Figure 5.3: Average SU Queue Curve

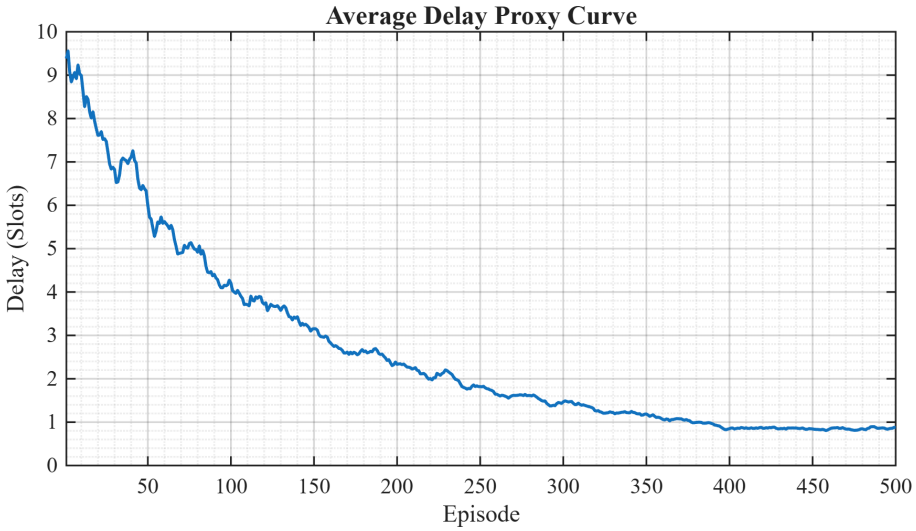


Figure 5.4: Average Delay Proxy Curve

5.1.4. Primary User Protection

PU protection metrics is the most crucial indicator of the system's adherence with the simualtion design. Throughout the course of the 500 episode run, the PU outage rate falls below the 10 db threshold; remaining at or close to zero as depicted in Figure 5.6.

The SU achieves throughput of nearly 80 Mbps and channel utilisation of 87%, demonstrating that it is aggressively exploiting available spectrum. the multi-threshold energy detection scheme that responds to uncertain channel conditions with reduced power transmission rather than full power access; and the interference penalty in the reward function that trains the DQL agent to associate channel access in high occupancy conditions with negative outcomes.

Over the course of the training run, the interference rate, the percentage of SU transmission slots increases from roughly 0.11 to roughly 0.27 in later episodes as shown in Figure 5.5. The primary outage is not always the result of an interference that may happen, the SU's interference power is lowered. The higher SU collision rate shows that the agent's willingness to learn, accessing channels and that the agent has implicitly learned.

The SU utilization inclines from roughly 0.38 in early training episodes to roughly 0.87 in the convergence rate as depicted in Figure 5.5. The other important result of the DQL earning process is this nearly doubling of the utilization rate. At 0.87 channel utilization, the secondary network is accessing the shared spectrum in 87% of simulation slots a level of intensity that would be impossible without intelligent channel selection.

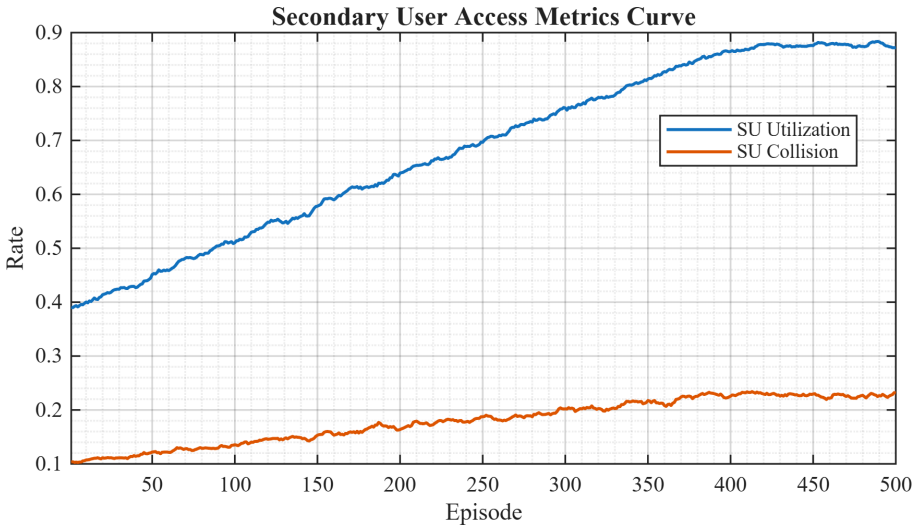


Figure 5.5: Secondary User Access Metrics Curve

The divergence between the interference event rate and the outage rate; the former inclines while the latter remains constant as in Figure 5.6 maybe the study's most important finding. By using the mechanisms to access uncertain channels frequently enough to achieve high secondary throughput while keeping the fraction of interference events that cause primary degradation practically insignificant, it demonstrates how the DQL framework can strike a balance between opportunistic aggressiveness and primary protection.

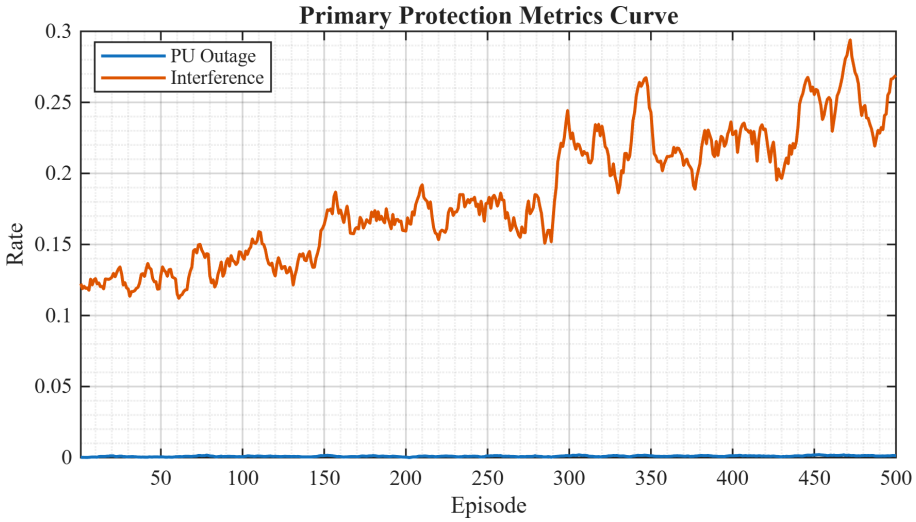


Figure 5.6: Primary Protection Metrics Curve

Figure 5.7 shows the worst case active PU SINR over training episodes. Even if the curve declines gradually,the SINR remained far above the PU Outage of 10 db showing that the PUs are protected while the SUs are trying to access available spectrum opportunities.

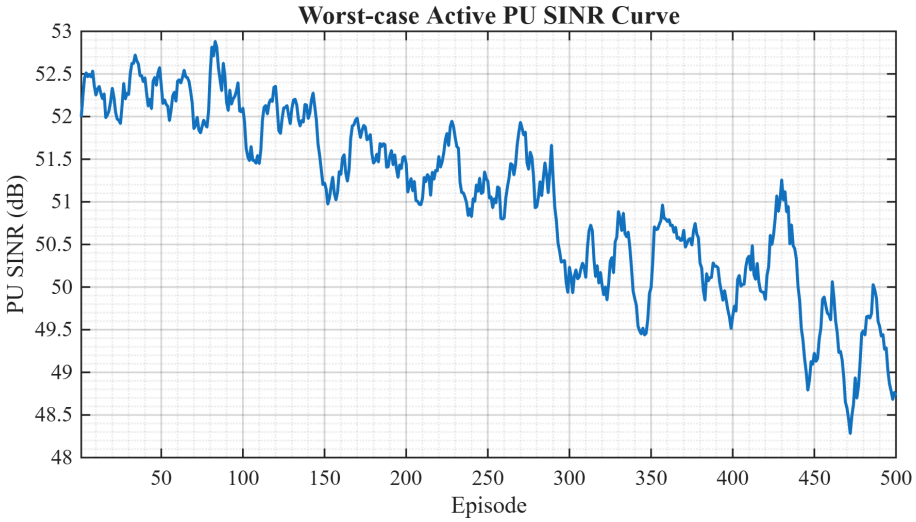


Figure 5.7: Worst Case Active PU SINR Curve

5.1.5. Spectrum Utilization and Idle Sensing Fraction

The idle sensing fractions, which happen when the SU access point senses a channel as idle but the agent decides not to transmit, show the DQL agent's learned risk management strategy. Since a very low idle sensing fraction would cause the agent to broadcast whenever it identifies an idle channel, inaccurate idle sensing could lead to significant interference rates, A high idle sensing percentage which would leave available spectrum unused, would be a sign of excessive conservatism.

As shown in Figure 5.8, the converged phase of training Episodes, the idle sensing fraction stabilizes at approximately 0.64 to 0.68. This figure indicates that about 64% to 68% of channel sensing observations are classified as idle. Despite the ongoing little fluctuations, the idle sensing fraction's consistency across episodes indicates the agent has adopted a constant risk management strategy rather than pursuing new access tactics.

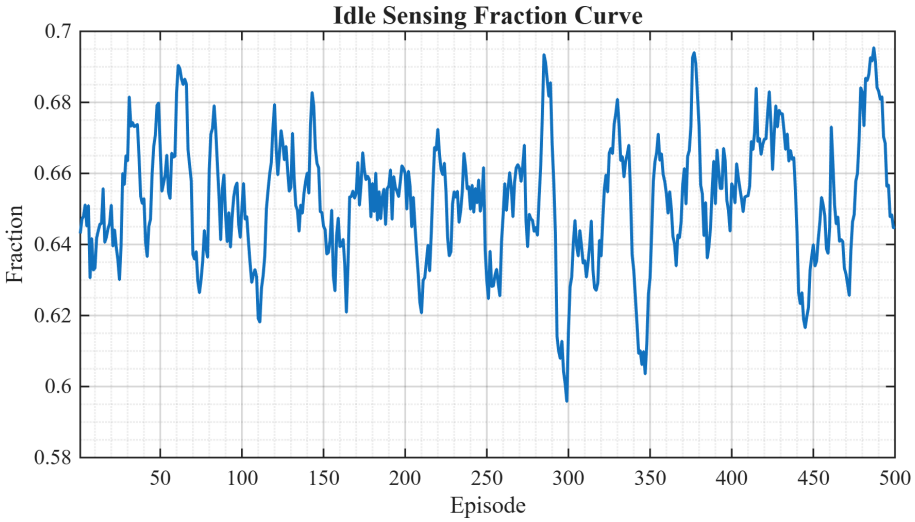


Figure 5.8: Idle Sensing Fraction Curve

Figure 5.9 shows that the PU outage remained low. It fluctuates due to the Markov traffic model applied, fading and interference conditions. In Figure 5.10 the curve appears to be gradually increasing from early training episodes and becomes stable around episode 450 showing that the DQL agent seems to be improving the efficiency of spectrum utilization.

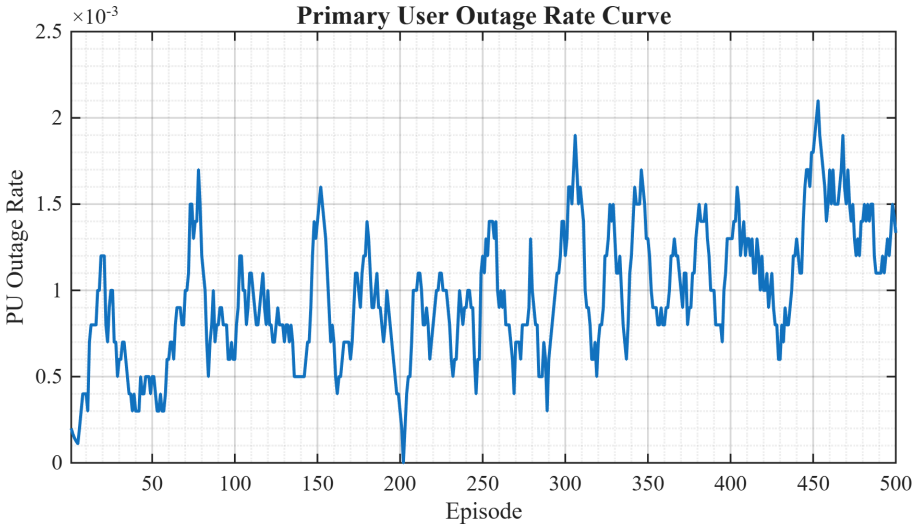


Figure 5.9: PU Outage Rate Curve

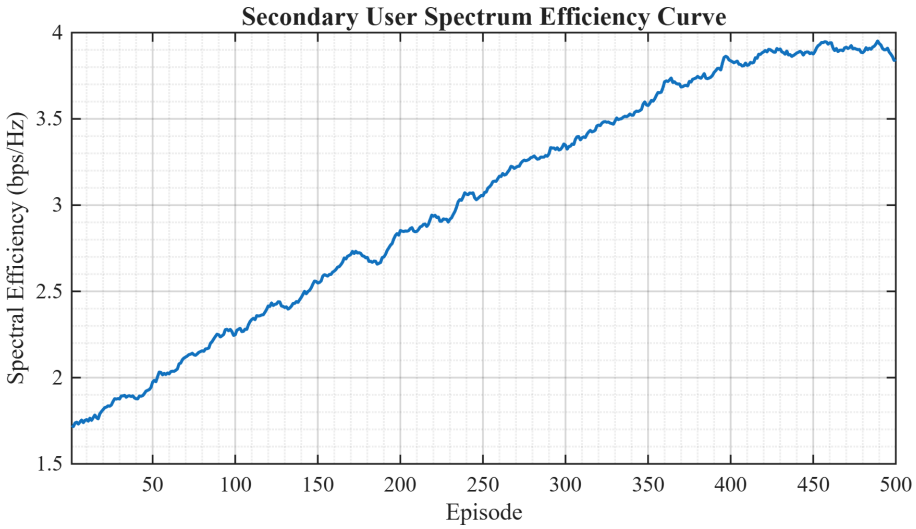


Figure 5.10: SU Spectrum Efficiency Curve

Figure 5.11 shows how reliable the PU protection is. The reliablity is close to 1 throughout the episodes. The fluctuation is mainly because of the changing PU traffic conditions, propagation and SU access attempt. Figure 5.12 presents the PU SINR margin. Although the curve gradually decreases it mainly stayed above the required threshold. The result shows that SU are accessing idle spectrum while keeping the PU priority protected.

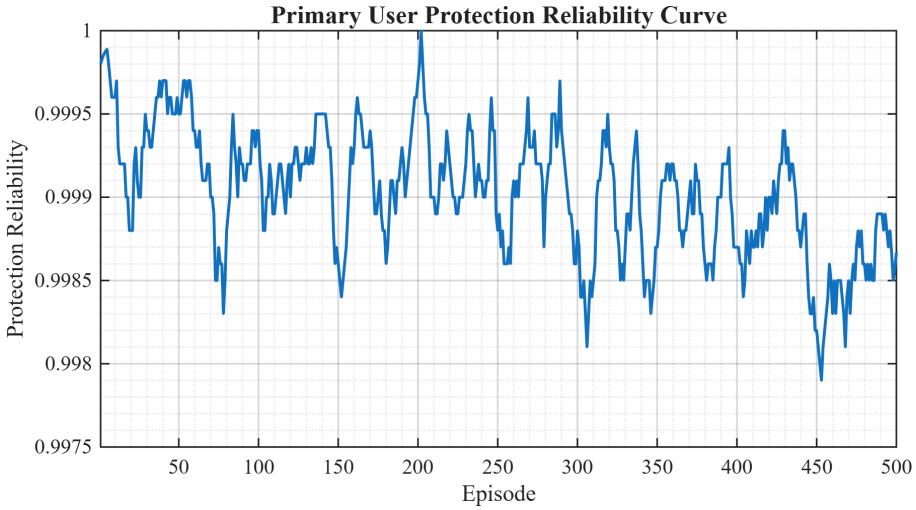


Figure 5.11: PU Protection Reliability Curve

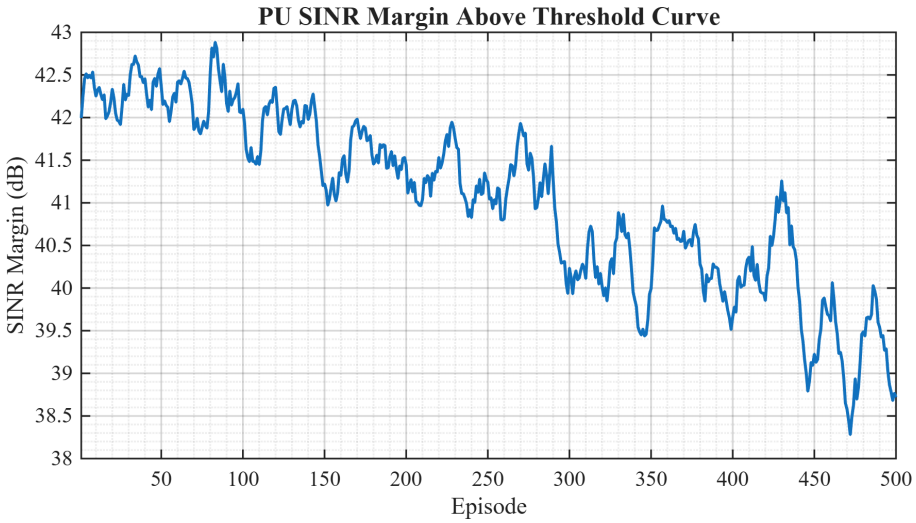


Figure 5.12: PU SINR Margin Curve

Figure 5.13 shows the cumulative average reward of the DQL agent during training. The figure shows a steady incline indicating the agent's improvement in long term effective decision making. It means the agent is learning to select better access actions that increase SU utilization while reducing interference to PUs.

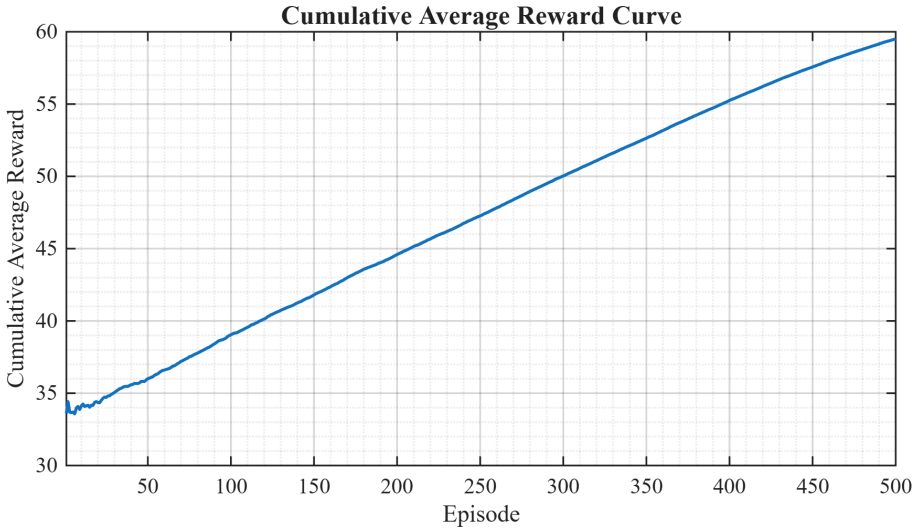


Figure 5.13: Cumulative Average Reward Curve

Figure 5.14 depicts a steady upward curve indicating the DQL agent is gradually improving its channel access policy and achieving better throughput as learning progresses. Figure 5.15 presents the epsilon greedy exploration decay curve. At the start of the training episode it is high around 1. It linearly decreases till episode 400 which means the agent is exploring many possible actions and stablizes from episode 400-500 meaning it has learned an access policy.

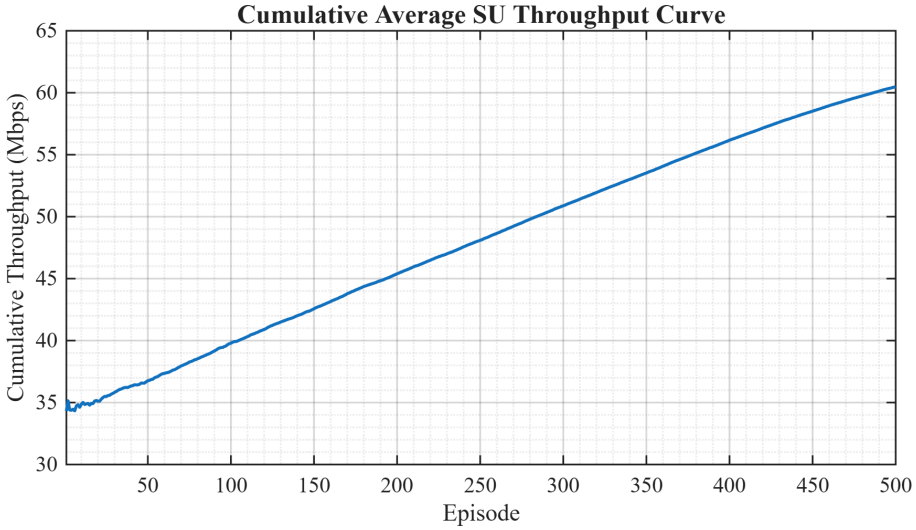


Figure 5.14: Cumulative Average SU Throughput Curve

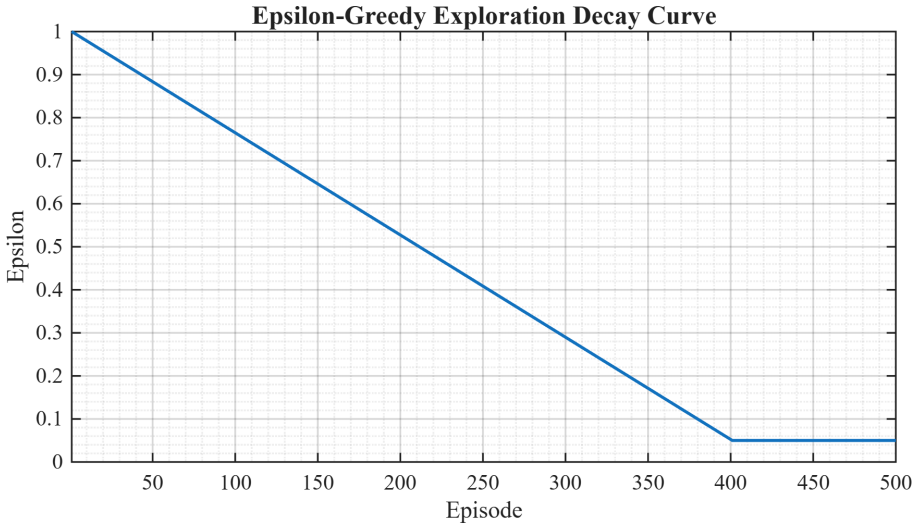


Figure 5.15: Epsilon Decay Curve

Figure 5.16 shows a close up review of the reward convergence. The reward fluctuates due to PU traffic randomness, propagation behavior and sensing uncertainty but it also remained relatively stable.

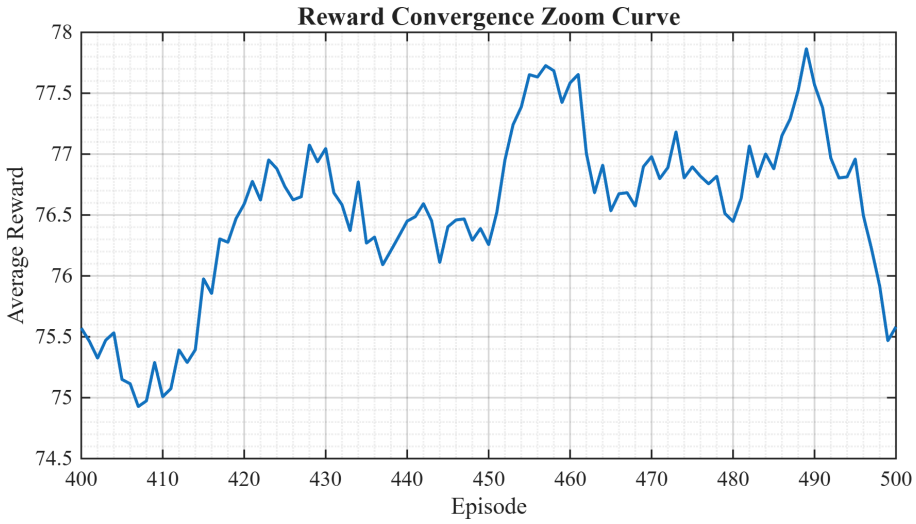


Figure 5.16: Reward Convergence Zoom Curve

Figure 5.17 depicts the trade off between SU throughput and PU outage rate. The results show the possibility of achieving high throughput while keeping the PU outage low. Figure 5.18 also shows the trade off between SU utilization and collision. The Figure shows that as SU utilization increase, the collision rate also increases which is believable but the good news is the collision rate is controlled.

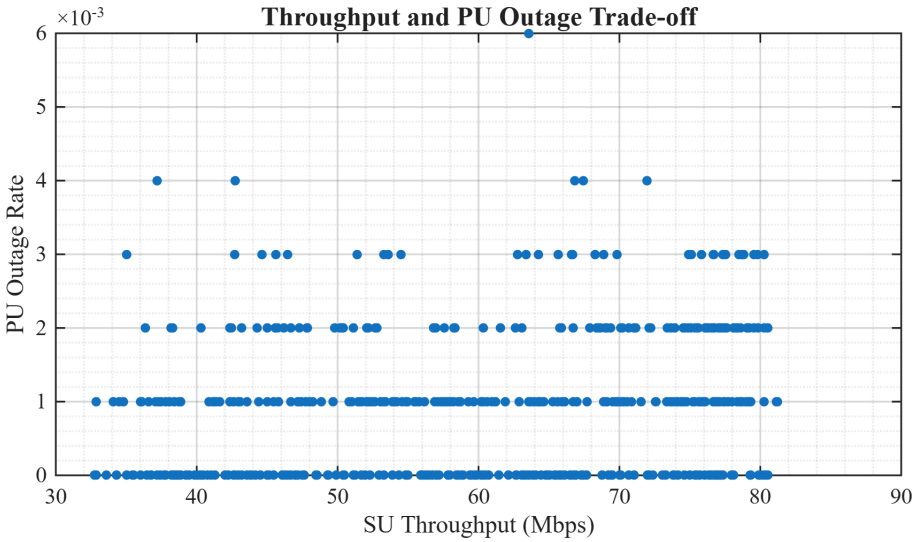


Figure 5.17: Throughput vs PU Outage Tradeoff

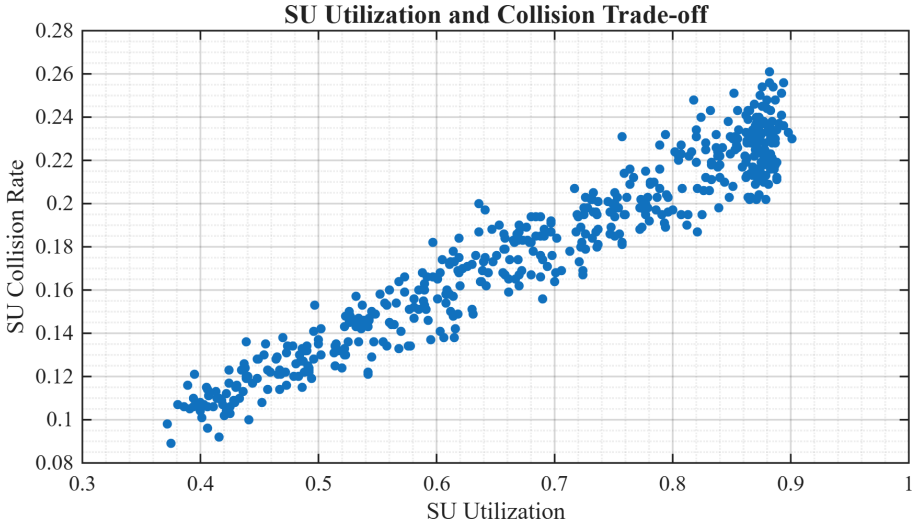


Figure 5.18: SU Utilization vs Collision Tradeoff

Figure 5.19 shows the relationship between queue and delay. As delay increases, the queue also increases since more waiting packets lead to longer transmission delay. It shows the importance of reducing SU queue backlog is crucial for improving delay. In Figure 5.20 the framework is summarized. The bar chart shows key metrics and their average value.Figure 5.21 depicts the summary of final PU protection and collision in the proposed DQL-based DSS frame work. The figure shows that PU outage remained low almost zero and PU's priority is also protected. The interference and SU collision are higher because the SU's active attempt to access channel but the lowered PU outage proved that the these values are controlled. Justifying that the DQL agent and LBT/CSMA-CA mechnism contributed in balancing PU protection with SU access.

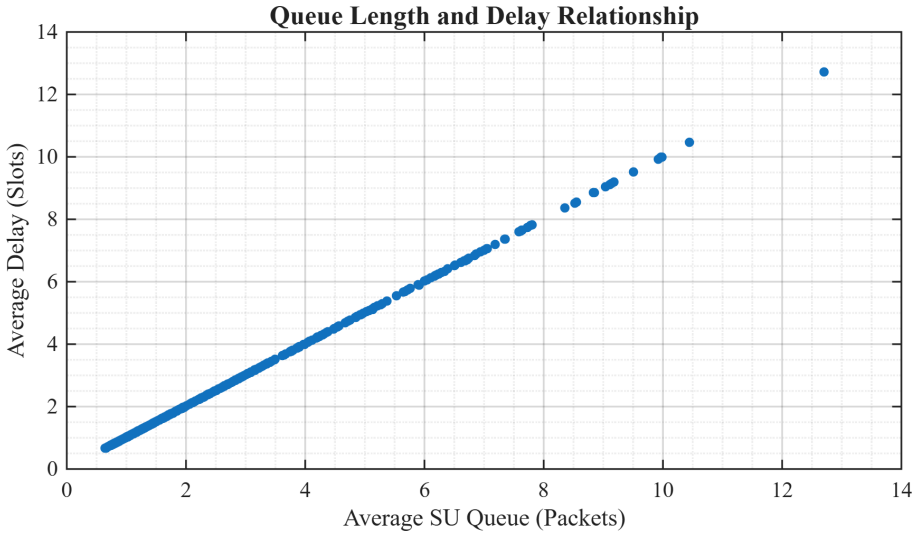


Figure 5.19: Queue vs Delay Relationship

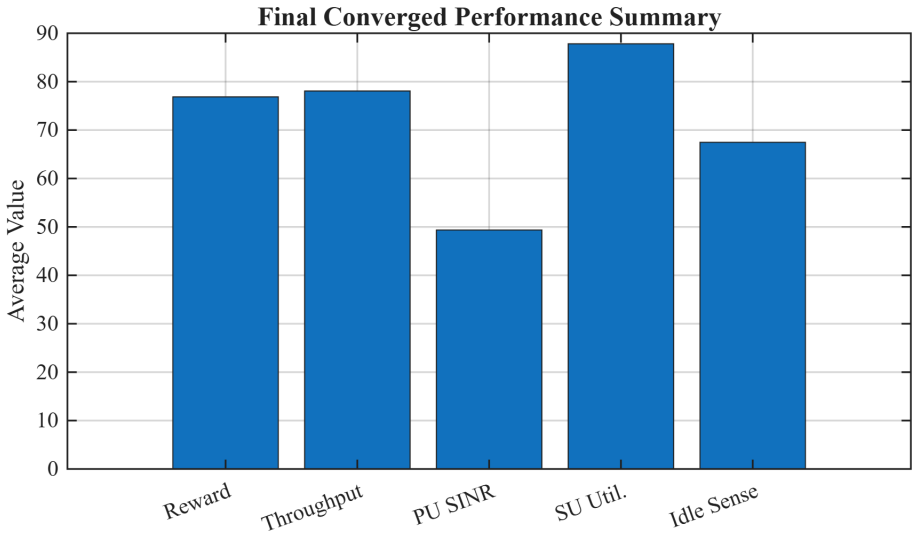


Figure 5.20: Final Performance Summary Bar Chart

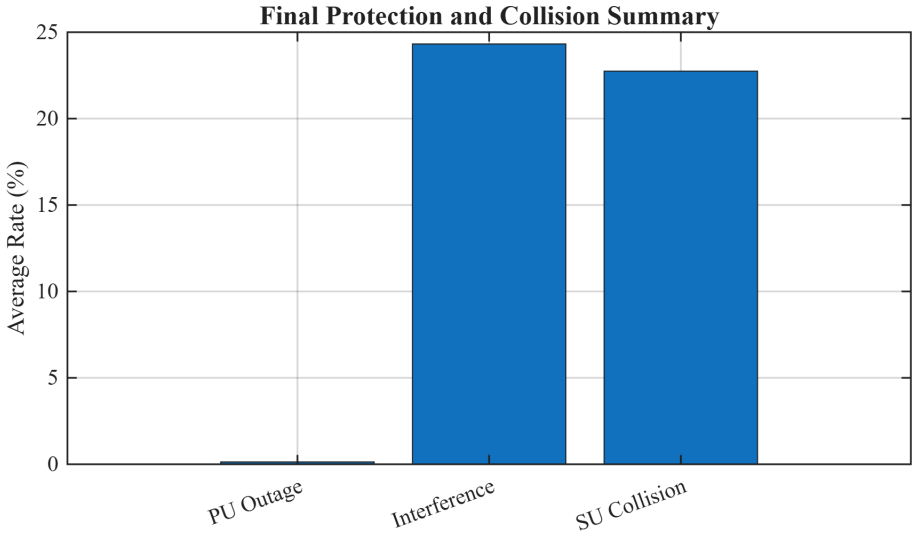


Figure 5.21: Final Protection Collision Summary

Chapter 6

Conclusion and Future Work

6.1. Conclusion

This thesis sets out to address the problem of spectral underutilization in licensed 5GNR spectrum using the DQL-based spectrum access framework. The work's consistent and coherent results support the thesis's main claim, which is that DQL provides a principled, flexible approach to DSS in high density 5G network environment. In order to enable SUs to opportunistically access idle channel while protecting PU's priority, the suggested system integrated multi-threshold energy detection, Markov chain-based traffic modeling. LBT/CSMA-CA access behavior and DQL-based decision making.

The results showed the DQL-agent improved reward convergence, improved SU throughput, spectrum usage and queue and delay behavior gradually learned effective access strategy keeping the PU outage low, demonstrating the DQL-based spectrum access provides an adaptive access methods in dynamic wireless network environments. Therefore, the proposed frame work supports the effectiveness of using DQL for intelligent CR-based spectrum sharing in licensed 5G NR bands

6.2. Limitations

Every simulation study has its constraints. The DQL controller's centralized architecture is the first drawback. In the current set up , a single agent at the access point makes all channel selection decisions for the ten SUs. The ohter drawback is the simplification of the propagation model the 3GPP bsed pathloss model used are still static even though they offer significantly more accurate depiction of the propagation environments than free space models. Third is the lack of hardware implementation-based experimental validation. All of the findings in this thesis are simulation, although the simulation settings are based on actual measurements and published standards, the gap between simulation and real world performance is an ever present concern in wireless communications research.

6.2.1. Future Work

The limitations and conclusions of the study point to several particular directions for future research. multi-agent DQL extensions are the most sensible and important direction for future study. By converting the single-agent design in to multi-agent configuration, where numerous SU access points each maintain their own DQL rules and collaborate through shared state information, the system would be able to grow to multi-cell secondary deployments. The primary research challenge in this expansion is controlling the non-stationarity caused by concurrently learning agents. Another way to advance the suggested framework from a simulation study to a production-ready implementation is through integration with ORAN infrastructure, deploying the DQL controller as an xAPP within the near-realtime RIC on ORAN architecture would be able to receive sensing data and issue channel access recommendations through standardized interfaces making it interoperable with a wide range of base stations and access point hardware; proving monitoring and management infrastructure needed to track system performance. Lastly, the increasingly importance of higher frequency bands in 5G and 6G would be addressed by extending to multi-band and mmWave spectrum sharing.

Bibliography

[1] I. Chochliouros, A. Spiliopoulou *et al.*, "Energy efficiency concerns and trends in future 5g network infrastructures. energies (mdpi) 14 (17), 5932 (2021)."

[2] M. K. Luka, O. U. Okereke, E. E. Omizegba, and E. C. Anene, "Blockchains for spectrum management in wireless networks: A survey," *arXiv preprint arXiv:2107.01005*, 2021.

[3] M. Tekanyi, J. A. Orokpo, H. Abdulkareem, and T. O. Alexander, "An improved channel selection algorithm (icsa) for spectrum hand-off in cognitive radio ad hoc network," *Indonesian Journal of Engineering and Science*, vol. 4, no. 1, pp. 001–011, 2023.

[4] A. Temtam and D. Popescu, "Cognitive radio spectrum sensing using tdsc: Evaluating performance in wimax and lte networks," *American Journal of Telecommunications*, vol. 1, no. 1, pp. 1–24, 2023.

[5] A. M. Al-Samman, T. Abd. Rahman, T. Al-Hadhrami, A. Daho, M. N. Hindia, M. H. Azmi, K. Dimyati, and M. Alazab, "Comparative study of indoor propagation model below and above 6 ghz for 5g wireless networks," *Electronics*, vol. 8, no. 1, p. 44, 2019.

[6] M. Aggarwal, T. Velmurugans, M. Karuppiah, M. M. Hassan, A. Almogren, and W. N. Ismail, "Probability-based centralized device for spectrum handoff in cognitive radio networks," *IEEE Access*, vol. 7, pp. 26 731–26 739, 2019.

[7] G. Pan, J. Li, and F. Lin, "A cognitive radio spectrum sensing method for an ofdm signal based on deep learning and cycle spectrum," *International Journal of Digital Multimedia Broadcasting*, vol. 2020, no. 1, p. 5069021, 2020.

[8] M. Girmay, V. Maglogiannis, D. Naudts, A. Shahid, and I. Moerman, "Coexistence scheme for uncoordinated lte and wifi networks using experience replay based q-learning," *Sensors*, vol. 21, no. 21, p. 6977, 2021.

[9] B. R. K. Kinoshita, "Spectrum sharing between cellular and wi-fi networks based on deep reinforcement learning."

[10] A. Orsino, G. Araniti, A. Molinaro, and A. Iera, "Effective rat selection approach for 5g dense wireless networks," in *2015 IEEE 81st vehicular technology conference (VTC Spring)*. IEEE, 2015, pp. 1–5.

[11] M. Moneesh, T. S. Tejaswi, T. S. Yeshwanth, M. Harshitha, and G. Chakravarthy, "Cooperative spectrum sensing using dqn in crn," *EAI Endorsed Trans. Mob. Commun. Appl.*, vol. 6, p. e4, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:236771458

[12] S. B. Shrote and S. Poshattiwar, "Dynamic spectrum sensing for 5g cognitive radio networks using optimization technique," *Journal of Electrical Systems*, vol. 20, no. 3s, pp. 1221–1231, 2024.

[13] J. Huang, Y. Yang, G. He, Y. Xiao, and J. Liu, "Deep reinforcement learning-based dynamic spectrum access for d2d communication underlay cellular networks," *IEEE Communications Letters*, vol. 25, no. 8, pp. 2614–2618, 2021.

[14] S. Szott, K. Kosek-Szott, P. Gawłowicz, J. T. Gómez, B. Bellalta, A. Zubow, and F. Dressler, "Wi-fi meets ml: A survey on improving ieee 802.11 performance with machine learning," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1843–1893, 2022.

[15] A. Doshi and J. G. Andrews, "Distributed proximal policy optimization for contention-based spectrum access," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021, pp. 340–344.

[16] ——, "Combining contention-based spectrum access and adaptive modulation using deep reinforcement learning," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2022, pp. 189–193.

[17] G. Frangulea, P. Assimakopoulos, S. Lagén, and B. Bojović, "Sensing-driven spatial reuse in shared spectrum: ns-3 implementation and evaluation of radiation null steering," in *Proceedings of the 2025 International Conference on ns-3*, 2025, pp. 11–19.

[18] M. Silva, J. Santos, and M. Curado, "The path towards virtualized wireless communications: A survey and research challenges," *Journal of Network and Systems Management*, vol. 32, no. 1, p. 12, 2024.

[19] S. Kim, "A new fair-efficient spectrum allocation scheme for the lte and wifi coexistence platform," *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, no. 1, p. 80, 2022.

[20] Y. Maruyama, K. Kinoshita, K. Murakami, and K. Kawano, "A spectrum sharing method based on adaptive threshold management between non-cooperative wimax/wifi providers," *ICN 2014*, p. 159, 2014.

[21] M. R. Fasihi and B. L. Mark, "Qos-aware state-augmented learnable framework for 5g nr-u/wi-fi coexistence: Impact of parameter selection and enhanced collision resolution," in *2026 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2026, pp. 590–595.

[22] Z. Zhong, J. Zhao, and C. Li, "Outdoor-to-indoor channel measurement and coverage analysis for 5g typical spectrums," *international Journal of Antennas and Propagation*, vol. 2019, no. 1, p. 3981678, 2019.

[23] U. Ali, G. Caso, L. De Nardis, K. Kousias, M. Rajiullah, Ö. Alay, M. Neri, A. Brunstrom, and M.-G. Di Benedetto, "Data-driven analysis of outdoor-to-indoor propagation for 5g mid-band operational networks," *Future Internet*, vol. 14, no. 8, p. 239, 2022.

[24] E. Ramos and A. Orsino, "Effects of beamforming and antenna configurations on mobility in 5g nr," *arXiv preprint arXiv:1811.06415*, 2018.

[25] A. Z. Hossain, N. B. Hassim, W. Hassan, W. A. Indra, S. G. Herawan, and M. Z. A. B. A. Aziz, "A planar 2× 2 mimo antenna array for 5g smartphones," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021.

[26] G. Geraci, A. Garcia-Rodriguez, D. López-Pérez, A. Bonfante, L. G. Giordano, and H. Claussen, "Operating massive mimo in unlicensed bands for enhanced coexistence and spatial reuse," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1282–1293, 2017.

[27] B. Hamdaoui, B. Khalfi, and N. Zorba, "Dynamic spectrum sharing in the age of millimeter-wave spectrum access," *IEEE Network*, vol. 34, no. 5, pp. 164–170, 2020.

Chapter A

Important MATLAB Codes

This appendix presents the main MATLAB code sections used to implement the proposed DQL-based dynamic spectrum sharing simulation framework. The selected code segments include parameter initialization, primary user traffic modeling, DQL configuration, spectrum sensing, state formation, action decoding, reward computation, environment update, and result plotting.

A.1. Simulation Parameter Definition

```
clear; clc; close all;
rng(7);  % Reproducibility

p = struct();

% Spectrum sensing parameters
p.sensingMethod         = 'Energy detection';
p.numSenseSamples       = 500;
p.sensingInterval_s     = 1e-3;
p.noiseFloor_dBm        = -100;
p.lambda                = [1.1 1.4 1.6];
p.noiseVarianceNorm     = 1;

% 5G NR primary user network parameters
p.NRband                = 'n40';
p.fcGHz                 = 2.35;
p.bandwidthHz           = 20e6;
p.K                     = 4;
p.numPU                 = 4;
p.slotTime_s            = 1e-3;
p.gNB_Tx_dBm            = 30;
p.gNB_AntGain_dBi       = 8;
p.PU_SINR_threshold_dB  = 10;

% Secondary user network parameters
p.numAP                 = 1;
p.numSU                 = 10;
p.SU_Tx_full_dBm        = 15;
p.SU_Tx_reduced_dBm     = 10;
p.CWmin                 = 15;
p.CWmax                 = 1023;
```

```
p.TXOP_normal_s         = 2e-3;
p.TXOP_short_s          = 1e-3;

% Simulation control
p.numEpisodes           = 500;
p.slotsPerEpisode       = 1000;
p.markovUpdateSlots     = 10;
```

Listing A.1: Main simulation and sensing parameters

A.2. Primary User Markov Traffic Model

```
% MATLAB index: 1 = Idle, 2 = Light, 3 = Medium, 4 = Heavy
% Active PU mapping: idle = 0, light = 1, medium = 2, heavy = 4
p.activePUsByState = [0 1 2 4];
p.markovStateNames = {'Idle','Light','Medium','Heavy'};

Pmarkov = [0.70 0.20 0.08 0.02;
           0.25 0.50 0.20 0.05;
           0.10 0.25 0.50 0.15;
           0.05 0.15 0.30 0.50];

function newState = sampleMarkov(currentState, P)
    cdf = cumsum(P(currentState,:));
    r = rand;
    newState = find(r <= cdf, 1, 'first');
end
```

Listing A.2: Four-state Markov chain primary traffic model

A.3. Deep Q-Learning Configuration

```
dql = struct();

dql.stateDescription  = 'Sensing state, SU queue, previous action,
    backoff, Markov load';
dql.actionDescription = 'Channel selection, transmit, re-sense, defer,
    silent';

dql.alpha             = 0.001;
dql.gamma             = 0.9;
dql.epsStart          = 1.0;
dql.epsEnd            = 0.05;
dql.epsDecayEp        = 400;
dql.replaySize        = 50000;
dql.batchSize         = 64;
dql.trainStart        = 1000;
dql.hidden1           = 64;
dql.hidden2           = 64;
dql.targetUpdateEvery = 500;

% State vector dimension:
% K sensing states + SU queue + previous action + backoff + Markov load
dql.stateDim = p.K + 4;

```

```
% For each channel, five access modes:
% 1 = TX full, 2 = TX reduced, 3 = Re-sense, 4 = Defer, 5 = Silent
dql.modesPerChannel = 5;
dql.numActions      = p.K * dql.modesPerChannel;
```

Listing A.3: DQL parameter configuration

A.4. Reward Function

```
% Reward weights
wThroughput = 1.0;
wOutage     = 12.0;
wInterf     = 3.0;
wCollision  = 2.0;
wDelay      = 0.02;

% Reward computation
reward = wThroughput * slotResult.suThroughputMbps ...
       - wOutage     * slotResult.puOutage ...
       - wInterf     * slotResult.interferenceEvent ...
       - wCollision  * slotResult.suCollision ...
       - wDelay      * suQueuePackets;
```

Listing A.4: Reward function used for DQL learning

A.5. Spectrum Sensing and Multi-Threshold Classification

```
function [senseClass, energyNorm, activePUchannels, activePUs] =
    senseEnvironment(p, markovState)

    activePUs = p.activePUsByState(markovState);

    activePUchannels = false(1,p.K);
    if activePUs > 0
        activePUchannels(1:activePUs) = true;
    end

    noiseLin = dbm2lin(p.noiseFloor_dBm);
    energyNorm = zeros(1,p.K);

    for k = 1:p.K
        % 500-sample energy estimate
        noiseSamples = noiseLin * (randn(p.numSenseSamples,1).^2);
        measured = mean(noiseSamples);

        if activePUchannels(k)
            d3D_gNB_AP = distance3D(p.gNB_to_building_2D_m, ...
                p.gNB_height_m, p.AP_height_m);

            PLout = pathlossUMi(d3D_gNB_AP, p.fcGHz);
            shadow = p.shadowNLOS_dB * randn;
            fadingGain_dB = fading_dB(p.rayleighFading);

            puRxAtAP_dBm = p.gNB_Tx_dBm + p.gNB_AntGain_dBi ...
                - PLout - p.buildingAtten_dB - shadow + fadingGain_dB;
```

```
28
29              measured = measured + dbm2lin(puRxAtAP_dBm);
30          end
31
32          energyNorm(k) = max(0.01, measured/noiseLin);
33      end
34
35      % 0 = idle, 1 = uncertain, 2 = likely PU, 3 = active PU
36      senseClass = zeros(1,p.K);
37
38      for k = 1:p.K
39          if energyNorm(k) < p.lambda(1)
40              senseClass(k) = 0;
41          elseif energyNorm(k) < p.lambda(2)
42              senseClass(k) = 1;
43          elseif energyNorm(k) < p.lambda(3)
44              senseClass(k) = 2;
45          else
46              senseClass(k) = 3;
47          end
48      end
49  end
```

Listing A.5: Energy detection and multi-threshold sensing

A.6. DQL State Formation

```
1  function s = buildState(senseClass, suQueuePackets, prevAction, ...
2                          backoffCounter, markovState, p, dql)
3
4      sSense = senseClass(:) / 3;
5      sQueue = min(suQueuePackets/100, 1);
6      sPrev  = prevAction / dql.numActions;
7      sBack  = min(backoffCounter/p.CWmax, 1);
8      sLoad  = (markovState-1) / 3;
9
10     s = [sSense; sQueue; sPrev; sBack; sLoad];
11 end
```

Listing A.6: DQL state vector formation

A.7. Action Decoding

```
1  function [ch, mode] = decodeAction(action, dql)
2
3      ch = ceil(action / dql.modesPerChannel);
4      mode = action - (ch-1) * dql.modesPerChannel;
5
6      % mode = 1: full-power transmission
7      % mode = 2: reduced-power transmission
8      % mode = 3: re-sense
9      % mode = 4: defer
10     % mode = 5: silent
11 end
```

Listing A.7: DQL action decoding into channel and access mode

A.8. Training Loop

```
for ep = 1:p.numEpisodes

    epsVal = max(dql.epsEnd, dql.epsStart - ...
        (dql.epsStart-dql.epsEnd)*(ep-1)/dql.epsDecayEp);

    markovState = 1;
    prevAction = 1;
    suQueuePackets = 20;
    backoffCounter = randi([0 p.CWmin]);

    for slot = 1:p.slotsPerEpisode

        if mod(slot-1, p.markovUpdateSlots) == 0
            markovState = sampleMarkov(markovState, Pmarkov);
        end

        [senseClass, energyNorm, activePUchannels, activePUs] = ...
            senseEnvironment(p, markovState);

        state = buildState(senseClass, suQueuePackets, prevAction, ...
            backoffCounter, markovState, p, dql);

        if rand < epsVal
            action = randi(dql.numActions);
        else
            qValues = forwardNet(net, state);
            [~, action] = max(qValues);
        end

        [chosenChannel, mode] = decodeAction(action, dql);

        [slotResult, suQueuePackets, backoffCounter] = stepEnvironment(
            ...
            p, SUpos, APpos, activePUchannels, activePUs, ...
            chosenChannel, mode, suQueuePackets, backoffCounter);

        reward = wThroughput*slotResult.suThroughputMbps ...
               - wOutage*slotResult.puOutage ...
               - wInterf*slotResult.interferenceEvent ...
               - wCollision*slotResult.suCollision ...
               - wDelay*suQueuePackets;

        prevAction = action;
    end
end
```

Listing A.8: Main DQL training loop

A.9. Result Figure Generation

```
plotFolder = 'Thesis_Style_Individual_Figures';

if ~exist(plotFolder, 'dir')
    mkdir(plotFolder);
end

episodes = 1:p.numEpisodes;
smoothWindow = 10;

rewardSmooth     = movmean(avgReward, smoothWindow);
throughputSmooth = movmean(avgSUThroughput, smoothWindow);
puSinrSmooth     = movmean(avgPUSINR, smoothWindow);
queueSmooth      = movmean(avgQueue, smoothWindow);
delaySmooth      = movmean(avgDelaySlots, smoothWindow);

makeThesisFigure(episodes, rewardSmooth, [], ...
    'DQL Reward Convergence Curve', 'Episode', 'Average Reward', ...
    {}, fullfile(plotFolder,'01_DQL_Reward_Convergence.png'));

makeThesisFigure(episodes, throughputSmooth, [], ...
    'Secondary User Throughput Curve', 'Episode', 'Throughput (Mbps)', ...
    {}, fullfile(plotFolder,'02_SU_Throughput_Curve.png'));

makeThesisFigure(episodes, puSinrSmooth, [], ...
    'Worst-case Active PU SINR Curve', 'Episode', 'PU SINR (dB)', ...
    {}, fullfile(plotFolder,'03_Worst_Case_PU_SINR_Curve.png'));
```

Listing A.9: Exporting individual thesis-style simulation figures

Bibliography

[1] I. Chochliouros, A. Spiliopoulou *et al.*, "Energy efficiency concerns and trends in future 5g network infrastructures. energies (mdpi) 14 (17), 5932 (2021)."

[2] M. K. Luka, O. U. Okereke, E. E. Omizegba, and E. C. Anene, "Blockchains for spectrum management in wireless networks: A survey," *arXiv preprint arXiv:2107.01005*, 2021.

[3] M. Tekanyi, J. A. Orokpo, H. Abdulkareem, and T. O. Alexander, "An improved channel selection algorithm (icsa) for spectrum hand-off in cognitive radio ad hoc network," *Indonesian Journal of Engineering and Science*, vol. 4, no. 1, pp. 001–011, 2023.

[4] A. Temtam and D. Popescu, "Cognitive radio spectrum sensing using tdsc: Evaluating performance in wimax and lte networks," *American Journal of Telecommunications*, vol. 1, no. 1, pp. 1–24, 2023.

[5] A. M. Al-Samman, T. Abd. Rahman, T. Al-Hadhrami, A. Daho, M. N. Hindia, M. H. Azmi, K. Dimyati, and M. Alazab, "Comparative study of indoor propagation model below and above 6 ghz for 5g wireless networks," *Electronics*, vol. 8, no. 1, p. 44, 2019.

[6] M. Aggarwal, T. Velmurugans, M. Karuppiah, M. M. Hassan, A. Almogren, and W. N. Ismail, "Probability-based centralized device for spectrum handoff in cognitive radio networks," *IEEE Access*, vol. 7, pp. 26 731–26 739, 2019.

[7] G. Pan, J. Li, and F. Lin, "A cognitive radio spectrum sensing method for an ofdm signal based on deep learning and cycle spectrum," *International Journal of Digital Multimedia Broadcasting*, vol. 2020, no. 1, p. 5069021, 2020.

[8] M. Girmay, V. Maglogiannis, D. Naudts, A. Shahid, and I. Moerman, "Coexistence scheme for uncoordinated lte and wifi networks using experience replay based q-learning," *Sensors*, vol. 21, no. 21, p. 6977, 2021.

[9] B. R. K. Kinoshita, "Spectrum sharing between cellular and wi-fi networks based on deep reinforcement learning."

[10] A. Orsino, G. Araniti, A. Molinaro, and A. Iera, "Effective rat selection approach for 5g dense wireless networks," in *2015 IEEE 81st vehicular technology conference (VTC Spring)*. IEEE, 2015, pp. 1–5.

[11] M. Moneesh, T. S. Tejaswi, T. S. Yeshwanth, M. Harshitha, and G. Chakravarthy, "Cooperative spectrum sensing using dqn in crn," *EAI Endorsed Trans. Mob. Commun. Appl.*, vol. 6, p. e4, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:236771458

[12] S. B. Shrote and S. Poshattiwar, "Dynamic spectrum sensing for 5g cognitive radio networks using optimization technique," *Journal of Electrical Systems*, vol. 20, no. 3s, pp. 1221–1231, 2024.

[13] J. Huang, Y. Yang, G. He, Y. Xiao, and J. Liu, "Deep reinforcement learning-based dynamic spectrum access for d2d communication underlay cellular networks," *IEEE Communications Letters*, vol. 25, no. 8, pp. 2614–2618, 2021.

[14] S. Szott, K. Kosek-Szott, P. Gawłowicz, J. T. Gómez, B. Bellalta, A. Zubow, and F. Dressler, "Wi-fi meets ml: A survey on improving ieee 802.11 performance with machine learning," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1843–1893, 2022.

[15] A. Doshi and J. G. Andrews, "Distributed proximal policy optimization for contention-based spectrum access," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021, pp. 340–344.

[16] ——, "Combining contention-based spectrum access and adaptive modulation using deep reinforcement learning," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2022, pp. 189–193.

[17] G. Frangulea, P. Assimakopoulos, S. Lagén, and B. Bojović, "Sensing-driven spatial reuse in shared spectrum: ns-3 implementation and evaluation of radiation null steering," in *Proceedings of the 2025 International Conference on ns-3*, 2025, pp. 11–19.

[18] M. Silva, J. Santos, and M. Curado, "The path towards virtualized wireless communications: A survey and research challenges," *Journal of Network and Systems Management*, vol. 32, no. 1, p. 12, 2024.

[19] S. Kim, "A new fair-efficient spectrum allocation scheme for the lte and wifi coexistence platform," *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, no. 1, p. 80, 2022.

[20] Y. Maruyama, K. Kinoshita, K. Murakami, and K. Kawano, "A spectrum sharing method based on adaptive threshold management between non-cooperative wimax/wifi providers," *ICN 2014*, p. 159, 2014.

[21] M. R. Fasihi and B. L. Mark, "Qos-aware state-augmented learnable framework for 5g nr-u/wi-fi coexistence: Impact of parameter selection and enhanced collision resolution," in *2026 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2026, pp. 590–595.

[22] Z. Zhong, J. Zhao, and C. Li, "Outdoor-to-indoor channel measurement and coverage analysis for 5g typical spectrums," *international Journal of Antennas and Propagation*, vol. 2019, no. 1, p. 3981678, 2019.

[23] U. Ali, G. Caso, L. De Nardis, K. Kousias, M. Rajiullah, Ö. Alay, M. Neri, A. Brunstrom, and M.-G. Di Benedetto, "Data-driven analysis of outdoor-to-indoor propagation for 5g mid-band operational networks," *Future Internet*, vol. 14, no. 8, p. 239, 2022.

[24] E. Ramos and A. Orsino, "Effects of beamforming and antenna configurations on mobility in 5g nr," *arXiv preprint arXiv:1811.06415*, 2018.

[25] A. Z. Hossain, N. B. Hassim, W. Hassan, W. A. Indra, S. G. Herawan, and M. Z. A. B. A. Aziz, "A planar 2× 2 mimo antenna array for 5g smartphones," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021.

[26] G. Geraci, A. Garcia-Rodriguez, D. López-Pérez, A. Bonfante, L. G. Giordano, and H. Claussen, "Operating massive mimo in unlicensed bands for enhanced coexistence and spatial reuse," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1282–1293, 2017.

[27] B. Hamdaoui, B. Khalfi, and N. Zorba, "Dynamic spectrum sharing in the age of millimeter-wave spectrum access," *IEEE Network*, vol. 34, no. 5, pp. 164–170, 2020.